



DEPARTAMENTO DE FÍSICA DE LA MATERIA CONDENSADA

TRABAJO DE FIN DE MÁSTER

---

# Propagación de epidemias, movilidad y estrategias de control

---

MÁSTER DE MODELIZACIÓN MATEMÁTICA, ESTADÍSTICA Y COMPUTACIÓN

*Autor*  
PABLO VALGAÑÓN RUIZ

*Supervisor*  
DR. JESÚS GÓMEZ GARDEÑES  
*Ponente*  
DR. ROBERTO BARRIO GIL

27 de junio de 2021

# Índice

<b>1. Resumen</b>	<b>1</b>
<b>2. Introducción</b>	<b>2</b>
<b>3. Marco teórico</b>	<b>4</b>
3.1. Redes Complejas . . . . .	4
3.1.1. Grafos y su representación . . . . .	4
3.1.2. Distribución de grado . . . . .	5
3.1.3. Sistemas en el mundo real . . . . .	6
3.1.4. Redes Erdős-Renyi . . . . .	7
3.1.5. Redes libres de escala . . . . .	7
3.2. Metapoblaciones . . . . .	8
3.2.1. Matriz Origen-Destino . . . . .	9
3.3. Modelos compartimentales . . . . .	10
3.3.1. Modelo SIS . . . . .	10
3.3.2. Modelo SIR . . . . .	11
3.3.3. Modelo SEAPIDR . . . . .	11
3.3.4. Diagramas de fase . . . . .	13
<b>4. Modelo de metapoblaciones con movilidad recurrente e individuos indistinguibles</b>	<b>14</b>
4.1. Verificación de las ecuaciones . . . . .	15
4.2. Cálculo del umbral epidémico . . . . .	17
4.3. Problemas y limitaciones del formalismo . . . . .	18
<b>5. Modelo de metapoblaciones con movilidad recurrente e individuos distinguibles</b>	<b>18</b>
5.1. Verificación de las ecuaciones . . . . .	19
5.2. Cálculo del umbral epidémico . . . . .	20
<b>6. Flujos críticos de movilidad humana</b>	<b>23</b>
<b>7. Desarrollando un modelo más complejo</b>	<b>26</b>
7.1. Densidad de población y función de contactos . . . . .	26
7.2. Procesos de reacción diurnos y nocturnos . . . . .	27
7.3. Modelo compartimental SEAPIDR . . . . .	28
7.3.1. El número básico de reproducción $R_0$ . . . . .	28
7.3.2. Calculando $R_0$ para el modelo compartimental SEAPIDR . . . . .	29
7.4. Implementación del modelo en el formalismo de individuos indistinguibles . . . . .	32
7.4.1. Verificación de las ecuaciones . . . . .	34
7.5. Implementación del modelo en el formalismo de individuos distinguibles . . . . .	35
7.5.1. Verificación de las ecuaciones . . . . .	36
7.6. Cálculo del umbral epidémico con metapoblaciones . . . . .	37
7.6.1. Linealización de las ecuaciones del formalismo indistinguible . . . . .	38
7.6.2. Linealización de las ecuaciones del formalismo distinguible . . . . .	39
<b>8. Flujos críticos para el diagnóstico de brotes epidémicos</b>	<b>42</b>
<b>9. Aplicación al caso del estado de Nueva York: diferentes escalas espaciales</b>	<b>43</b>
<b>10. Conclusiones</b>	<b>46</b>

## 1. Resumen

El trabajo trata de la modelización de epidemias y procesos de contagio en poblaciones humanas teniendo en cuenta la movilidad de personas entre regiones. En las secciones 3 y 4 se presenta el marco existente en el campo de redes complejas, metapoblaciones y epidemias, explicando el formalismo del que partimos para el desarrollo de los modelos a lo largo del trabajo. Introducimos el llamado formalismo de individuos distinguibles, explicando las características que lo diferencian de su predecesor. Veremos cómo obtener ecuaciones deterministas que describen la evolución del sistema y su verificación, además de expresiones a partir de las mismas que nos ayudarán a ver el efecto de la movilidad en la capacidad que tienen los brotes epidémicos de propagarse o extinguirse. En la sección 6 mostramos una aplicación del nuevo formalismo para la toma de medidas de control, basada en la restricción de ciertos flujos de movilidad humana entre poblaciones, y cómo se puede mitigar el impacto de una epidemia con eficiencia siguiendo el modelo. Después añadiremos una serie de mejoras importantes al modelo para ambos formalismos en la sección 7, donde también se verá cómo estas modificaciones alteran los resultados. En la sección 8 mostraremos cómo utilizar el formalismo para centrar los tests y diagnosticar casos de forma más eficiente. Para finalizar, estudiamos el efecto que la escala de la resolución de los datos utilizados tiene en el impacto del modelo, que demuestra la importancia del uso de redes complejas para el estudio de estos fenómenos.

The paper deals with the modeling of epidemics and contagion processes in human populations taking into account the mobility of people between regions. The existing framework in the field of complex networks, metapopulations and epidemics is presented in the sections 3 and 4, explaining the formalism from which we start for the development of the models throughout the paper. We introduce the so-called formalism of distinguishable individuals, explaining the characteristics that differentiate it from its predecessor. We will see how to obtain deterministic equations describing the evolution of the system and how to verify them, as well as deriving the expressions that will help us see the effect of mobility on the ability of epidemic outbreaks to spread or die out. In section 6 we show an application of the new formalism for taking control measures, based on the restriction of certain human mobility flows between populations, and how the impact of an epidemic can be efficiently mitigated by following the model. We will then add a number of important improvements to the model for both formalisms in section 7, where we will also see how these modifications alter the results. In section 8 we will show how to use the formalism to focus tests and diagnose cases more efficiently. Finally, we study the effect that the scale of the resolution of the data has on the impact of our model, which demonstrates the importance of using complex networks to study these phenomena.

## 2. Introducción

La Física de Sistemas Complejos estudia las propiedades y comportamientos que surgen a partir de las interacciones entre las componentes de un sistema [1]. Para considerarse complejo, éste debe constar de una gran cantidad de agentes que se relacionan e interaccionan entre ellos sin la intervención de un controlador central. La razón por la que reciben el nombre de *complejos* se debe a la emergencia de propiedades que surgen a un nivel superior, es decir, pueden ser caracterizadas por variables macroscópicas, que no pueden explicarse directamente a partir de las acciones de componentes individuales, sino que vendrán determinados por la topología de la red que describe las relaciones entre ellos. Por esta razón, una de las disciplinas en auge es la Mecánica Estadística de Redes Complejas, que ha encontrado aplicaciones en diversas áreas de todos los rincones de la investigación científica. Campos de estudio como fenómenos de sincronización [2], congestiones de tráfico [3], ideas y rumores [4] o economía [5] se han beneficiado enormemente de las aplicaciones de redes complejas para simular y predecir la dinámica de los procesos característicos de estos sistemas.

La razón por la que recurrimos a la Física de Sistemas Complejos en todas estas disciplinas es que, a la hora de utilizar métodos convencionales, la alta no linealidad de las ecuaciones, unida a la gran relevancia de la estructura de interacciones, limitan el grado de precisión de aproximaciones habituales en otras disciplinas como las teorías de campo medio. Gracias al aumento de la capacidad de computación de los ordenadores en las últimas décadas, podemos explicar este tipo de fenómenos a escalas cada vez mayores y obtener mejores resultados, y es por ello que este campo está creciendo con tanta velocidad. Sin embargo, aunque sus aplicaciones sean tan extensas en la actualidad, la teoría de redes y grafos lleva siglos siendo desarrollada, desde que Euler publicó el primer trabajo de grafos en 1736 para resolver el famoso problema de los puentes de Königsberg [6].

En este trabajo se utilizará la teoría de redes complejas para estudiar la propagación de epidemias. Durante las últimas décadas se ha hecho evidente el creciente riesgo bajo el que nos encontramos de ser afectados severamente por este fenómeno. Mientras que las condiciones sanitarias en la mayor parte del mundo están mejorando, existen otros mecanismos cuyo efecto es igual de importante en cuanto a hacernos susceptibles a cualquier enfermedad, como el aumento de la densidad de población en núcleos urbanos o la cantidad de personas que viajan a diario a escala internacional. En este sentido, ambos fenómenos han dado lugar a la rápida expansión global de diferentes enfermedades cuyos primeros brotes estaban espacialmente localizados tales como el Zika, el Ébola o más recientemente la COVID-19. Precisamente, el nuevo coronavirus constituye la prueba más evidente de las dificultades que supone frenar la dispersión de patógenos con una infectividad media-alta en una sociedad tan conectada y móvil como la actual. Por este motivo, en las últimas décadas, los modelos matemáticos [7] han adquirido una gran relevancia ya que permiten predecir el avance de un brote y evaluar el impacto de diferentes políticas de control para mitigarlo. Una comprensión de este fenómeno y sus causas es crucial a la hora de responder a diversas situaciones de manera informada y conociendo las consecuencias de dichas políticas.

Es importante reconocer la importancia que tiene la movilidad humana a la hora de estudiar los procesos de contagio y difusión de una epidemia, y este es el concepto en el que se centra este trabajo. El campo de estudio del movimiento de las personas está recibiendo una cantidad enorme de nuevo contenido, ya sea debido a publicaciones científicas que tratan de predecir y simular estos patrones de viajes o debido al enorme flujo de datos que somos capaces de recopilar gracias a los avances tecnológicos (seguimiento de recorridos de billetes individuales [8], telefonía móvil, GPS [9], redes sociales) y a los esfuerzos por parte de gobiernos regionales (principalmente con censos y encuestas).

El contagio de enfermedades infecciosas es un fenómeno complicado que depende de una cantidad enorme de variables, lo que hace difícil encontrar una forma de estudiarlo y describirlo que sea tan simple como para que podamos extraer resultados analíticos, pero lo suficientemente elaborado como para dotar al modelo de un nivel de complejidad que permita capturar con precisión escenarios epidémico reales. Es por esta razón que es tan complicado encontrar un modelo que cumpla todas estas condiciones, y este será el principal objetivo del trabajo: partir de modelos ya existentes, encontrar puntos débiles y suposiciones que puedan ser incorrectas, y corregirlos añadiendo niveles de complejidad intentando mantener la mayor cantidad de generalidad posible.

La forma más común de abordar la modelización de epidemias es el uso de modelos compartimentales, que fueron introducidos en el siglo XX [10]. En ellos, se hace una aproximación de campo medio en una población y se asigna a los residentes diferentes estados, llamados compartimentos, que describen en qué fase de una infección se encuentran. Sin embargo, a pesar de ser muy efectivos al describir de forma cualitativa la evolución de una epidemia, no se ajustan con precisión a casos reales, donde los diferentes niveles de interacciones y movilidad se hacen evidentes y afectan enormemente a los resultados. La primera forma en la que se introdujeron las redes complejas para este caso fue al eliminar la aproximación de campo medio y formar una red de relaciones entre agentes, pudiendo eliminar la homogeneidad y estudiar casos específicos en los que ciertos individuos tengan muchos más contactos que otros, formen puentes entre cúmulos de otros agentes, etcétera. Mientras que esto supone una mejora, es evidente que no es suficiente si queremos tener en cuenta la movilidad de los agentes y el hecho de que nuestras relaciones con otros cambian constantemente.

Para introducir el efecto de la movilidad, se utilizan las metapoblaciones, que fueron introducidas originariamente en el campo de la ecología para estudiar la coexistencia de diferentes especies en el mismo ecosistema. Mediante el uso de metapoblaciones, modelizaremos los procesos de reacción-difusión que gobiernan la propagación de epidemias y comprobaremos cómo diferentes asunciones en torno a la naturaleza y escalas de los movimientos humanos pueden afectar a los resultados obtenidos y al impacto predecido de una epidemia, utilizando las herramientas proporcionadas por la teoría de redes complejas.

### 3. Marco teórico

#### 3.1. Redes Complejas

Una red [11] es, en términos generales, cualquier sistema que admite una representación matemática abstracta como un **grafo** cuyos nodos, o vértices, se identifican como los elementos del sistema, donde el conjunto de enlaces o conexiones representan la presencia de una relación o interacción entre ellos.

##### 3.1.1. Grafos y su representación

Para empezar con los conceptos más básicos, un grafo **no dirigido** consta de dos conjuntos  $G = (v, \epsilon)$ , donde  $v$  es un conjunto contable no nulo de elementos, que se corresponden con los vértices, y  $\epsilon$  es un conjunto de pares **no ordenados** de distintos vértices, formando los enlaces. El enlace  $(i, j)$  une los vértices  $i$  y  $j$ , que decimos están conectados. El tamaño de un grafo viene dado por el número de vértices  $N$  que lo componen, o en su lugar del número de enlaces  $L$ . Para un grafo no dirigido de  $N$  vértices, la cantidad máxima de enlaces posible es de  $L = \binom{N}{2}$ .

Por otro lado, un grafo **dirigido** consta de, además del conjunto de vértices, un conjunto de pares **ordenados**. En los grafos no dirigidos, una línea uniendo dos vértices los conecta en ambos sentidos, mientras que en los dirigidos puede conectarlos en cualquier sentido, o en los dos.

Desde un punto de vista matemático conviene definir un grafo a través de la **matriz de adyacencia**, que es una matriz  $N \times N$  definida tal que:

$$M_{ij} = \begin{cases} 1 & \text{si } (i, j) \text{ conectados} \\ 0 & \text{si } (i, j) \text{ no conectados} \end{cases}$$

En la figura 3.1 se muestran ejemplos de grafos y sus matrices de adyacencia asociadas.

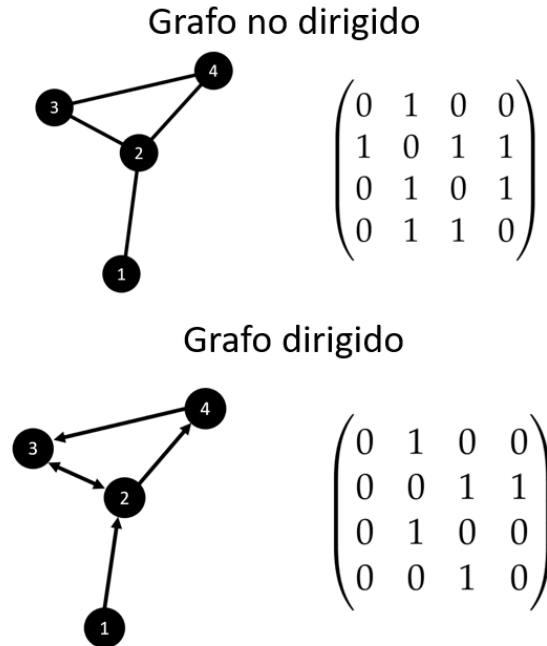


Figura 3.1

Notar que los grafos no dirigidos tienen matrices de adyacencia simétricas, por lo que estamos almacenando información repetida.

Otra característica de ciertas redes es que pueden ser pesadas. Esto significa que no todos los enlaces tienen la misma importancia, y los que tengan un valor nulo implican que no existe un enlace. Gracias a redes pesadas, podemos representar muchas más clases de relaciones entre vértices. Un ejemplo, que tiene relación con la aplicación en que se centra este trabajo, es la cantidad de personas que viajan de una ciudad a otra diariamente.

Es muy importante, a la hora de trabajar con grafos, tener en cuenta su dispersión (*sparseness*), o lo vacíos que se encuentran. Las redes pueden tener un amplio rango de enlaces por lo que a la hora de almacenar y trabajar con un grafo nos interesa saber cómo escala su cantidad de enlaces en función de sus vértices. Si  $L \sim N^\alpha$ , con  $\alpha < 2$ , decimos que es disperso. Es decir, conforme aumentamos el valor de  $N$ , la cantidad  $L/N^2$  va a tender a cero, y la matriz de adyacencia estará compuesta en casi su totalidad por elementos nulos. Por otro lado, si  $\alpha = 2$ , el grafo es denso.

Cuando trabajamos con grafos esparcidos, la matriz de adyacencia supone un gasto innecesario de memoria, ya que casi todas las entradas son ceros, por lo que usamos otras formas de representar la red. Una solución son las **listas de adyacencia**, dadas por  $l(i, v \in \nu(i))$ , donde  $\nu(i)$  es el conjunto de vecinos del nodo  $i$ . Por ejemplo, la lista de adyacencia correspondiente al grafo no dirigido de la figura 3.1 es:

$$\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \begin{bmatrix} 2 & & & \\ 1 & 3 & 4 & \\ 2 & 4 & & \\ 2 & 3 & & \end{bmatrix}$$

Si la red fuera pesada, bastaría con añadir al lado los pesos de cada enlace. Esto usará entonces el doble de memoria por enlace no nulo que la matriz de adyacencia, pero se compensa con todos los nulos que no hay que almacenar.

El ahorro de memoria no es la única ventaja de trabajar con listas de adyacencia. Muchas operaciones que llevemos a cabo en la red serán además mucho más rápidas. Por ejemplo, si tenemos que sumar funciones dependientes de los vecinos del nodo  $i$ , tal que la función es cero si el enlace no existe, tenemos dos posibilidades:

$$\sum_{j=1}^N f(M_{ij}) \quad \sum_{j \in \nu(i)} f(M_{ij})$$

Si hiciésemos esto para todos los nodos, veríamos que la primera forma llevaría un coste de  $N^2$  pasos, mientras que la segunda llevaría  $L$  pasos, siendo esta cantidad mucho menor en una red dispersa.

### 3.1.2. Distribución de grado

El grado  $k$  de un vértice  $i$  se define como el número de enlaces del grafo que comparte con otros nodos. En el caso de redes dirigidas distinguimos los grados de entrada  $k_{in}$  y salida  $k_{out}$ , y el grado de un nodo será la suma de ambos.

$$k_{in,i} = \sum_j M_{ji} \quad k_{out,i} = \sum_j M_{ij}$$

Conocer el grado de cada nodo nos permite hacernos una idea de su relevancia, ya que en el estudio de la propagación de una epidemia los nodos de mayor grado tendrán un impacto mayor. Cuando tengamos una red pesada, podemos además definir la fuerza  $s$  de cada nodo, de

la misma manera que el grado, dado que la matriz de adyacencia incluye los pesos. Cuando los pesos son independientes de la topología, la fuerza de un nodo crecerá linealmente con su grado ( $s \approx \langle w \rangle k$ ), donde  $\langle w \rangle$  es el peso medio de las conexiones, y  $k$  es el grado del nodo.

Una de las formas más comunes de describir una red es a través de su **distribución de grado**, que se define como la probabilidad  $P(k)$  de que cualquier vértice que elijamos tenga grado  $k$ . Se puede construir a partir del histograma normalizado del grado de los nodos de la red. Con esto, podemos conocer los momentos dados por

$$\langle k^n \rangle = \sum_{k=1}^{\infty} k^n P(k),$$

y la media de una red no dirigida viene dada por:

$$\langle k \rangle = \sum_{k=1}^{\infty} k P(k) = \frac{1}{N} \sum_i k_i = \frac{2L}{N}.$$

En una red dispersa,  $\langle k \rangle \ll N$ . Por ejemplo, en el modelo de Ising en dos dimensiones, donde cada nodo de la red tiene un enlace con sus cuatro vecinos, el valor de  $\langle k \rangle$  se mantiene constante aunque aumentemos el valor de  $N$  indefinidamente.

Hay muchas otras formas de describir una red (*betweenness*, *closeness*, *clustering*), pero en la aplicación de este trabajo la distribución de grado es la más relevante, especialmente a la hora de crear grafos *realistas*.

### 3.1.3. Sistemas en el mundo real

Con los avances en tecnología e información hemos obtenido acceso a una enorme cantidad de redes a gran escala, a través de datos de movilidad (transporte, comercio, viajes), redes sociales, entre otras, que tienen características muy particulares. En concreto, casi todas estas redes son **libres de escala** [12]. Esto es indicación de una gran heterogeneidad entre los nodos de la red. Unos pocos, llamados *Hubs*, acumulan una gran cantidad de enlaces, mientras que el resto tendrán muy pocos en comparación. Las distribuciones de grado de estas redes tienen cola pesada (*heavy tail*).

Por ejemplo, en una red homogénea, la distribución de grado suele ser normal o Poissoniana, que son de cola ligera, mientras que una red heterogénea de distribución en ley de potencias es de cola pesada.

En concreto, para redes cuya distribución de grado siga una ley de potencias dada por  $P(k) = k^{-\gamma}$ , tal que  $2 < \gamma \leq 3$ , parece que la media del grado está delimitada. Por simplicidad, consideramos la distribución  $P(k)$  en todos los reales para calcular la integral

$$\int_1^{k_c} k P(k) dk = \int_1^{k_c} k^{1-\gamma} dk = \frac{1}{2-\gamma} (k_c^{2-\gamma} - 1). \quad (3.1)$$

Vemos que la media es estable, incluso cuando el tamaño de la red crece indefinidamente (y  $k_c \rightarrow \infty$ ). Sin embargo, si queremos estimar el error que cometemos al asumir que el grado de un vértice es  $\langle k \rangle$ , nos vamos al segundo momento

$$\int_1^{k_c} k^2 P(k) dk = \int_1^{k_c} k^{2-\gamma} dk = \frac{1}{3-\gamma} (k_c^{3-\gamma} - 1) \quad (3.2)$$

el cual diverge con el tamaño de la red. Es entonces cuando nos damos cuenta de que la media del grado no nos da ninguna información sobre el grado de cada nodo, y tenemos una red libre



de escala. Para  $\gamma < 2$  ni siquiera la media converge.

Vista esta propiedad de la mayoría de sistemas que observamos en el mundo real, nos planteamos formas de crear una red que tenga esta característica.

### 3.1.4. Redes Erdős-Renyi

Una forma de construir redes aleatorias homogéneas, o redes de Erdős-Renyi [13], es partiendo de un número de nodos  $N$  y, para cada par de nodos, decidir aleatoriamente si crear un enlace entre ellos o no, con probabilidad  $p$ . La distribución del grado  $k$  entonces será una binomial dada por la expresión

$$\binom{N-1}{k} p^k (1-p)^{N-1-k}$$

y se trata de una distribución de cola ligera, con una media de  $\langle k \rangle = p \cdot (N-1)$ . Al aumentar el tamaño de la red, la distribución tenderá a una poissoniana, donde todos los nodos tienen un grado similar. Es por lo tanto un ejemplo de red muy homogénea.

### 3.1.5. Redes libres de escala

Para simular el comportamiento de redes reales, Baràbasi y Albert introdujeron una forma de crear redes [14] de forma que, en vez de trabajar con un número fijo de nodos  $N$ , se van introduciendo vértices uno por uno y la red crece de forma dinámica.

El método consiste en empezar con una pequeña red aleatoria, que se puede generar por otros métodos como el de Erdős-Renyi, y a cada paso se introduce un nuevo vértice, que tendrá una posibilidad de crear un enlace con cada nodo de la red proporcional al grado de dicho nodo. De esta forma, va a haber una preferencia por los nodos de mayor grado, y conforme crece la red y los hubs aumentan sus enlaces, esa desigualdad se va amplificando. Queda una distribución de ley de potencias  $k^{-3}$ .

En el caso que nos concierne en este trabajo, donde las redes representan la movilidad entre barrios, ciudades o provincias, basaremos nuestro análisis en redes de movilidad reales extraídas de datos oficiales de la oficina de censos de Estados Unidos, con lo que no será necesario crear redes sintéticas <sup>1</sup>.

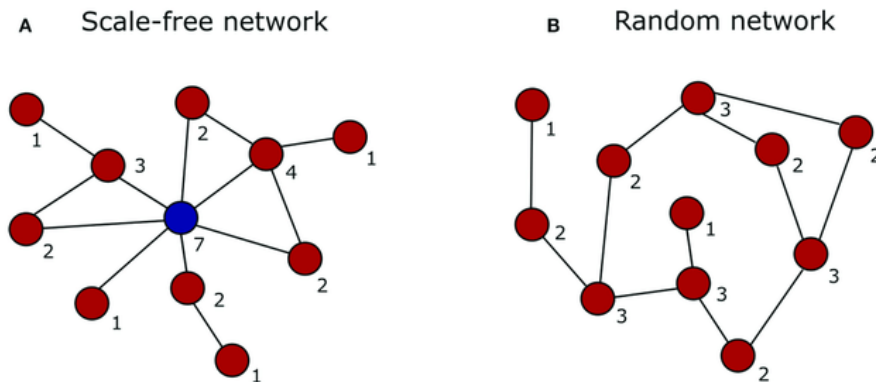


Figura 3.2: Representación de una red libre de escala frente a una aleatoria (Erdős-Renyi).

Fuente: ResearchGate

<sup>1</sup>Fuente de datos de movilidad de ciudades de Estados Unidos: <https://lehd.ces.census.gov/data/lodes/LODES7/>

Una vez hemos visto los conceptos más importantes sobre las redes, vamos a ver cómo lo aplicamos a los modelos de movilidad.

### 3.2. Metapoblaciones

Un modelo de propagación de epidemias debe describir la forma en la que los individuos se relacionan y tienen contactos entre ellos. Una primera forma de simular estos sistemas es con una red entre individuos, donde los enlaces representan los contactos. Sin embargo, esto no tiene en cuenta la difusión que se observa en la vida real: las personas no se relacionan siempre con los mismos individuos. Aquí es donde entra el papel de las metapoblaciones.

Un modelo de metapoblaciones describe la forma en la que los agentes interactúan en un sistema, a partir de describir su movimiento entre diferentes localizaciones espaciales. Una metapoblación es una red compleja compuesta de subpoblaciones, que son conjuntos de personas reunidas en un lugar y que interactúan entre ellas. Estas subpoblaciones serán tratadas como los nodos o vértices de una red compleja, y los enlaces representarán la movilidad entre las mismas. Por ejemplo, dentro de una ciudad, nuestra red puede describir la cantidad de personas que viajan del barrio en el que residen a otro donde esté localizado su lugar de trabajo.

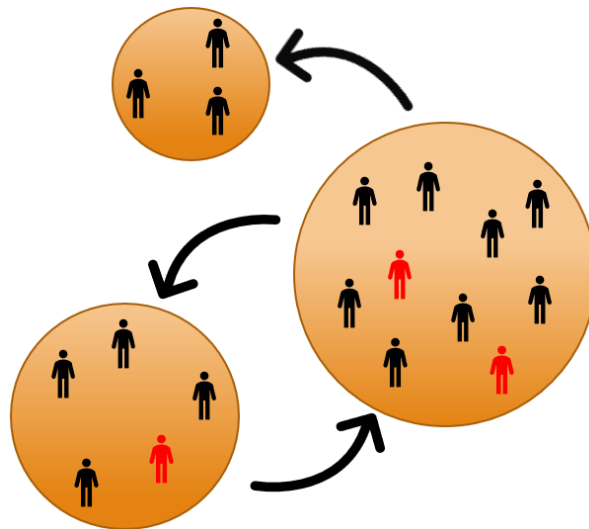


Figura 3.3: Ejemplo de metapoblación

Las metapoblaciones han sido utilizadas extensivamente en el estudio de epidemias, sin embargo aún queda por resolver la cuestión de cuál es la mejor forma de parametrizar modelos de movilidad, y cómo esa elección puede afectar a los resultados que obtenemos. Con la reciente abundancia de datos de todo tipo, desde censos hasta seguimiento por GPS, es difícil interpretar toda esta información y aplicarla a un caso concreto. Dependiendo del origen de estos datos, podemos encontrar información a muchas escalas diferentes, tanto espaciales como temporales. Podemos estudiar epidemias fijándonos en los cambios demográficos con los pasos de los años (migraciones, generaciones de nuevos portadores de la enfermedad) o en desplazamientos diarios entre barrios dentro de una ciudad. Juzgar cuál es el formalismo más apropiado para cada caso es un aspecto al que se le ha dado menos importancia.

Las dos clases de modelos de movilidad más utilizadas son la **Euleriana** y la **Lagrangiana** [15, 16], donde las diferencias estructurales entre ambas afectan considerablemente a los resultados [17]. Los modelos de movimiento Euleriano no siguen el comportamiento de agentes individuales, sino que únicamente describen la tasa con la que personas de una subpoblación se desplazan a otra, sin que un agente en concreto recuerde su posición anterior a haberse movido. Por otro lado, un modelo de movimiento Lagrangiano incluye a cualquiera que describa la cantidad de tiempo que los agentes pasan en otros parches antes de volver a su lugar de origen, así que poseen memoria temporal y permiten introducir movimientos recurrentes.

Un ejemplo de movilidad Euleriana es el modelo de flujo, que describe los viajes como una difusión de la población de un lugar a otro. Si la cantidad de personas que se encuentran en un nodo  $i$  es  $N_i$ , se puede describir el cambio de las poblaciones en tiempo continuo como

$$\frac{dN_i}{dt} = - \sum_{j=1}^K f_{ij} N_i + \sum_{j=1}^K f_{ji} N_j \quad (3.3)$$

donde la población total de la metapoblación se mantiene constante con el tiempo, pero la de cada parche puede variar hasta que no se alcance un estado estacionario. En un modelo de portadores de una enfermedad, la difusión de los infectados se describirá también con esta ecuación.

Por otro lado, el modelo del viaje simple es un ejemplo de movilidad Lagrangiana. En él se diferencian los residentes de una subpoblación de los visitantes, los cuales regresarán a su nodo eventualmente. Se puede asociar a los agentes a un lugar de residencia de forma permanente, y sus ecuaciones de la dinámica en tiempo continuo son

$$\frac{dN_{ii}}{dt} = - \sum_{j=1}^K \phi_{ij} N_{ii} + \sum_{j=1}^K \tau_{ij} N_{ij}, \quad (3.4)$$

$$\frac{dN_{ij}}{dt} = - \tau_{ij} N_{ij} + \phi_{ij} N_{ii}. \quad (3.5)$$

En este modelo entonces una persona que se ha desplazado a otro parche debe volver a su residencia antes de poder moverse a otro distinto.

Estos modelos utilizan los parámetros  $f_{ij}$ ,  $\phi_{ij}$  y  $\tau_{ij}$  para representar los viajes. Vamos a introducir la notación que usaremos a lo largo del trabajo con el mismo objetivo y las propiedades que deben cumplir.

### 3.2.1. Matriz Origen-Destino

Llamamos matriz origen-destino (OD) a la matriz de adyacencia cuyos pesos representan los viajes entre subpoblaciones dentro de una metapoblación. Tenemos varias formas de almacenar esta información, dependiendo de cómo vamos a trabajar con ella.

Una forma puede consistir en almacenar en un vector la cantidad de personas  $n_i$  que reside en cada población, y la matriz OD representa el porcentaje de gente que se desplaza a otros nodos, dado como  $R_{ij}$ . Obviamente debe cumplir

$$\sum_j R_{ij} = 1. \quad (3.6)$$

Una segunda forma consiste en simplemente representar en la matriz OD la cantidad total de personas que se mueven entre nodos  $n_{ij}$ . Se relaciona con el formalismo anterior con las igualdades

$$n_i \cdot R_{ij} = n_{ij}, \quad (3.7)$$

$$\sum_j n_{ij} = n_i. \quad (3.8)$$

En todos los modelos que describiremos a lo largo del trabajo, distinguimos dos procesos: **reacción** y **difusión**. Este trabajo se centrará en desarrollar y estudiar un formalismo que explique y describa correctamente la difusión en una población, y para ello hemos introducido la teoría de redes complejas y las metapoblaciones. Sin embargo, no podemos probar correctamente los resultados sin introducir antes modelos que expliquen el proceso de reacción. Para ello, usaremos los ya existentes modelos compartimentales.

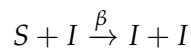
### 3.3. Modelos compartimentales

Los modelos compartimentales describen un número de estados en los que se puede encontrar un individuo, y cómo ocurre la transición de un estado a otro. Por ejemplo, dos estados que podemos encontrar respecto a cualquier enfermedad son los de susceptible e infectado, donde un individuo que se encuentra sano puede infectarse al tener un contacto con otro que ya haya contraído la enfermedad previamente. De esta forma, conociendo los contactos que ocurren entre los individuos, podemos simular la dinámica de su estado con el tiempo. Vamos a ver algunos modelos compartimentales relevantes a continuación.

#### 3.3.1. Modelo SIS

En este modelo los dos estados posibles son el de susceptible (S) e infectado (I). La dinámica se define de la siguiente forma:

Un individuo susceptible  $S$ , al entrar en contacto con un infectado, tiene una probabilidad  $\beta$  de contraer la enfermedad y pasar al estado  $I$ .



Un individuo infectado, en cada paso temporal discreto, tendrá una probabilidad de recuperarse y volver a ser susceptible de  $\mu$ .



Este es uno de los modelos más simples, y es utilizado para simular enfermedades de las que no se obtiene inmunidad tras contraerlas. Una de sus características es que la dinámica de un sistema en este modelo alcanza un estado estacionario, donde el número de personas que se recuperan a cada paso temporal es el mismo que el de individuos que se infectan.

Aprovechamos para introducir el **Acercamiento Macroscópico de Cadena de Markov** (MMCA). Los procesos de contagio tienen una naturaleza estocástica, donde cada individuo se puede encontrar cambiando de estado de forma aleatoria, así que es imposible predecir el estado microscópico del sistema en cualquier momento. Aun así, podemos conocer el estado macroscópico más probable utilizando ecuaciones deterministas que obedezcan una dinámica markoviana en tiempo discreto (el estado en el siguiente paso temporal solamente depende del anterior).

El estado macroscópico en este caso estará dado por la cantidad de personas (o el porcentaje) de una población que se encuentran en cada estado, y en el caso del modelo SIS evoluciona como

$$\rho^I(t+1) = (1-\mu)\rho^I(t) + (1-\rho^I(t))\Pi(t) \quad (3.9)$$

donde  $\rho^I(t)$  es la proporción de infectados de una población en el momento  $t$ , y  $\Pi(t)$  es la probabilidad de que un susceptible se contagie, que dependerá de los contactos que tenga con otros infectados. En general, si la probabilidad de contraer la enfermedad al tener un contacto con una persona infecciosa es  $\beta$ , contamos la cantidad de personas  $k$  con las que tendrá contacto y calculamos la probabilidad de que no se infecte con ninguna de ellas, para después obtener su probabilidad complementaria

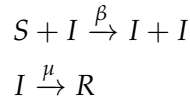
$$\Pi(t) = 1 - \left(1 - \beta\rho^I(t)\right)^k. \quad (3.10)$$

Por otro lado, la variable  $\rho^S(t)$  está restringida debido a que se debe cumplir

$$\rho^S(t) + \rho^I(t) = 1 \quad \forall t. \quad (3.11)$$

### 3.3.2. Modelo SIR

A este modelo se le añade un nuevo estado, que es el de recuperado (R). Funciona de la misma forma que el SIS, con la diferencia de que una persona que se haya recuperado de la enfermedad no va a volver a contraerla.



En la vida real, esto se da cuando el enfermo desarrolla inmunidad o incluso en el caso de defunción. Al contrario que en el modelo anterior, no existe un estado estacionario en el que haya personas infectadas, sino que la epidemia va a llegar a su fin en algún momento por la falta de individuos susceptibles. Las ecuaciones markovianas que describen la dinámica en este modelo son:

$$\rho^I(t+1) = (1-\mu)\rho^I(t) + (1-\rho^I(t) - \rho^R(t))\Pi(t), \quad (3.12)$$

$$\rho^R(t+1) = \rho^R(t) + \mu\rho^I(t). \quad (3.13)$$

Debido a que no se llega a un estado estacionario, es más difícil probar y comparar este modelo con simulaciones de agentes, así que durante el desarrollo de los distintos modelos del trabajo usaremos más a menudo el SIS.

### 3.3.3. Modelo SEAPIDR

Si queremos aplicar los modelos compartimentales a enfermedades reales, debemos incluir esquemas más complejos. En particular, en el caso de la COVID-19, el modelo SEAPIDR es un buen candidato para capturar sus características fundamentales. A continuación se detallan las propiedades de cada compartimento:

- Susceptible (S): mismo significado que en los anteriores modelos. Cuando entra en contacto con un infectado, tendrá una probabilidad de pasar al estado expuesto (E). La probabilidad depende del estado del agente con el que interactúa.
- Expuesto (E): ha contraído la enfermedad, pero aún no muestra síntomas ni es contagioso. En cada paso de tiempo, tiene una probabilidad  $(1-x)\nu$  de convertirse en infeccioso asintomático (A) y  $x\nu$  de pasar a ser infeccioso presintomático (P).

- Asintomático (A): no muestra síntomas pero tiene la capacidad de contagiar a un susceptible, con probabilidad  $\beta_A$ . A su vez, tiene una probabilidad  $\mu_A$  de recuperarse (R).
- Presintomático (P): como el asintomático, pero tiene una probabilidad distinta  $\beta_P$  de contagiar a un susceptible, y pasará a ser sintomático (I) con probabilidad  $\alpha$ .
- Infeccioso sintomático (I): puede contagiar a un susceptible con probabilidad  $\beta_I$ , tiene una tasa de recuperación (R) de  $\mu_I$ , pero al mostrar síntomas se puede hacer un test y ser detectado (D) con probabilidad  $\delta$ .
- Detectado (D): sigue estando enfermo pero no contagiará a nadie por cumplir cuarentena. Tiene una probabilidad  $\gamma$  de recuperarse (R).
- Recuperado (R): es el estado final de cualquier individuo infectado. Los recuperados no pueden volver a contraer la enfermedad.

Las ecuaciones que describen la dinámica del sistema serán las siguientes:

$$\begin{aligned}
\rho^E(t+1) &= (1 - \eta)\rho^E(t) + (1 - \rho^E(t) - \rho^A(t) - \rho^P(t) - \rho^I(t) - \rho^D(t) - \rho^R(t))\Pi(t), \\
\rho^A(t+1) &= (1 - \mu_A)\rho^A(t) + (1 - x)\eta\rho^E(t), \\
\rho^P(t+1) &= (1 - \alpha)\rho^P(t) + x\eta\rho^E(t), \\
\rho^I(t+1) &= (1 - \delta - \mu_I)\rho^I(t) + \alpha\rho^P(t), \\
\rho^D(t+1) &= (1 - \gamma)\rho^D(t) + \delta\rho^I(t), \\
\rho^R(t+1) &= \mu_A\rho^A(t) + \mu_I\rho^I(t) + \gamma\rho^D(t), \\
\Pi(t) &= 1 - \left(1 - \beta_A\rho^A(t) - \beta_P\rho^P(t) - \beta_I\rho^I(t)\right)^k.
\end{aligned} \tag{3.14}$$

En este trabajo no se tienen en cuenta las vacunas, ni la pérdida de inmunidad a largo plazo, pero se pueden incorporar en este modelo compartimental fácilmente, añadiendo formas de pasar de susceptible a recuperado y viceversa. Además, el compartimento de detectados (D) hace posible el estudio de políticas de confinamiento en función de los casos notificados. Es por lo tanto un modelo completo y suficientemente sencillo para caracterizar la evolución de la COVID-19.

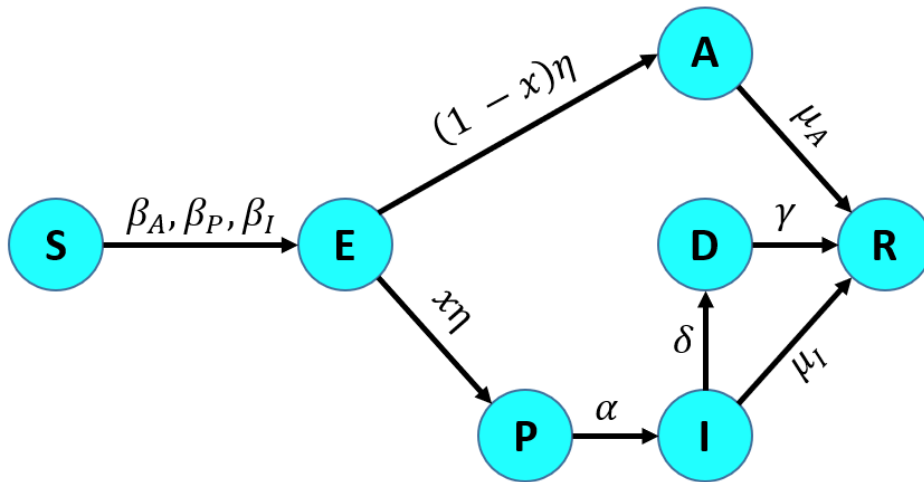


Figura 3.4: Diagrama de los compartimentos del modelo SEAPIDR y sus probabilidades de transición entre estados.

### 3.3.4. Diagramas de fase

Uno de los objetivos más importantes en el estudio de la propagación de epidemias es estimar el impacto que tendrá en la población. En el caso SIS, podemos considerar el impacto como el número de personas infectadas en el estado estacionario, y en un SIR podría ser la cantidad de recuperados al final de la simulación. En cualquier caso, ese número dependerá de la red y las características de la enfermedad, siendo el parámetro de mayor importancia la probabilidad de contagio por contacto  $\beta$ . Uno de los diagramas que más utilizamos será el que relaciona el impacto de la epidemia con  $\beta$ , como es el caso de la figura 3.5.

Existen dos fases separadas por el umbral epidémico  $\beta_c$ , que puede definirse como el valor de la probabilidad de contagio mínima necesaria para que exista un estado estacionario con un número de infectados distinto de cero. Por debajo de dicho umbral, los contagios no son suficientes y al llegar a cero infectados se extingue la epidemia.

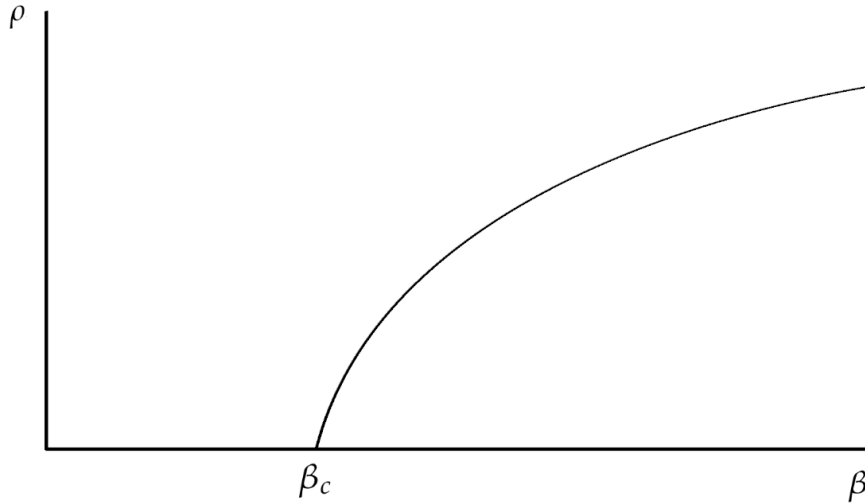


Figura 3.5: Diagrama de fases cualitativo para una epidemia. En el eje vertical se representa el porcentaje de infectados, y en el eje horizontal la probabilidad de contagio.  $\beta_c$  es el umbral epidémico.

En la siguiente parte veremos un modelo de metapoblaciones anterior que será el punto de partida de los desarrollos en este trabajo. Nos servirá como ejemplo para explicar otros conceptos y métodos básicos que usaremos a lo largo del trabajo, no sólo para desarrollar otros modelos en el formalismo de ecuaciones markovianas, sino también para ver cómo podemos comprobar su exactitud y extraer resultados analíticos de ellas.

## 4. Modelo de metapoblaciones con movilidad recurrente e individuos indistinguibles

Los modelos de **reacción-difusión** han sido estudiados en profundidad para el análisis de propagación de epidemias, ya que se puede pensar en ellas como el cambio de concentración de diferentes elementos, o sustancias, que en este caso es la cantidad de infectados en una población. El proceso de reacción es local y se refiere al cambio de unos elementos a otros, estando dado por las interacciones entre ellos, mientras que la difusión es la que causa que esas sustancias se mezclen ocupando el espacio disponible.

Mientras que existen muchos modelos de metapoblaciones que describen este tipo de procesos, un objetivo principal ha sido encontrar un formalismo en el que podamos complementar las simulaciones microscópicas de agentes individuales (que ocurren en tiempo discreto, y se basan en probabilidades tanto de viajes como de contagios y recuperaciones) con ecuaciones mesoscópicas y deterministas, que puedan predecir el resultado de las simulaciones anteriores. Con este objetivo, se propone un modelo en [18] que explica los contactos a nivel individual, con probabilidades de contagio  $\beta$  y recuperación  $\mu$ , con ecuaciones deterministas a partir de las cuales se pueden obtener resultados analíticos interesantes.

El nuevo modelo por lo tanto asignará a cada agente un nodo (o subpoblación) de residencia permanente, desde el que en cada paso temporal decidirá aleatoriamente un lugar de destino (difusión). Una vez hecho esto, se lleva a cabo el proceso de contagio, donde se considera que todas las personas dentro de un nodo tienen contacto entre ellas, y finalmente vuelven a sus lugares de origen. Esto se repite en cada paso temporal y se observa evolucionar el sistema. Es por lo tanto un modelo de movilidad Lagrangiana, ya que hace un seguimiento de los individuos conociendo su residencia de origen. De esta forma, podemos definir el estado mesoscópico del sistema como los porcentajes de personas que se encuentren en distintos compartimentos para cada nodo  $i$ . Por ejemplo, en el modelo SIR necesitaremos dos vectores  $\rho_i^I, \rho_i^R$ , de tamaño  $N$  igual al número de parches o nodos en la red, para describir el sistema por completo.

Como hemos dicho, este formalismo permite estimar la dinámica del estado mesoscópico mediante ecuaciones de Markov, que describimos a continuación para un modelo SIS. La primera ecuación es equivalente a la vista anteriormente (3.9), haciendo campo medio en cada parche de la metapoblación (consideramos a todos los individuos de una subpoblación **indistinguibles**).

$$\rho_i^I(t+1) = (1-\mu)\rho_i^I(t) + (1-\rho_i^I(t))\Pi_i(t) \quad \forall i, \quad 1 \leq i \leq N. \quad (4.1)$$

Para calcular  $\Pi_i(t)$  tenemos que introducir una nueva variable  $p$  que nos da la probabilidad de que una persona decida viajar de su residencia a un destino determinado. Si  $p = 0$ , todas las personas se quedan en su parche original, y si  $p = 1$  todos deciden moverse. Entonces  $p \cdot R_{ij}$  nos dará la probabilidad de que un individuo del parche  $i$  se encuentre en  $j$  cuando el contagio tenga lugar. La expresión para obtener la probabilidad de que una persona del nodo  $i$  se contagie es

$$\Pi_i(t) = (1-p)P_i(t) + p \sum_{j=1}^N R_{ij}P_j(t) \quad (4.2)$$

siendo  $P_i$  la probabilidad de que una persona que se encuentre en el nodo  $i$  en el momento de la reacción contraiga la enfermedad. Este proceso se llevará a cabo como hemos visto en (3.10), cada persona entrará en contacto con todas las demás que se encuentren en ese momento en el mismo parche, y podemos conocer la probabilidad de que no se contagie con ningún contacto.



$P_i(t)$  será la probabilidad complementaria del suceso anterior, por lo que

$$P_i(t) = 1 - \prod_{j=1}^N \left(1 - \beta \rho_j^I(t)\right)^{n_{i \rightarrow j}} \quad (4.3)$$

y solamente queda calcular cuántas personas del nodo  $i$  viajan al nodo  $j$ , dado por

$$n_{i \rightarrow j} = \delta_{ij}(1 - p)n_i + pR_{ji}n_j. \quad (4.4)$$

#### 4.1. Verificación de las ecuaciones

Las ecuaciones markovianas nos permiten estudiar las propiedades del modelo de forma analítica y rápida, pero primero tenemos que comprobar que sean correctas. Para ello, las ponemos a prueba frente a simulaciones de agentes basadas en algoritmos de **Monte Carlo**. En ellas, representamos cada uno de los individuos y el estado en el que se encuentran, de forma que no tendremos un proceso determinista, como es el caso de las ecuaciones de Markov. En su lugar, tenemos un sistema estocástico donde la localización y los cambios de estado se deciden aleatoriamente en cada paso temporal, de acuerdo con las reglas microscópicas establecidas para el modelo.

En la figura 4.1 vemos dos ejemplos de simulaciones para distintos valores de la probabilidad de contagio, donde representamos el impacto de la epidemia que medimos como la fracción total de infectados en la metapoblación, y viene dada por

$$\rho^I(t) = \frac{\sum_{j=1}^N n_j \rho_j^I(t)}{\sum_{j=1}^N n_j}. \quad (4.5)$$

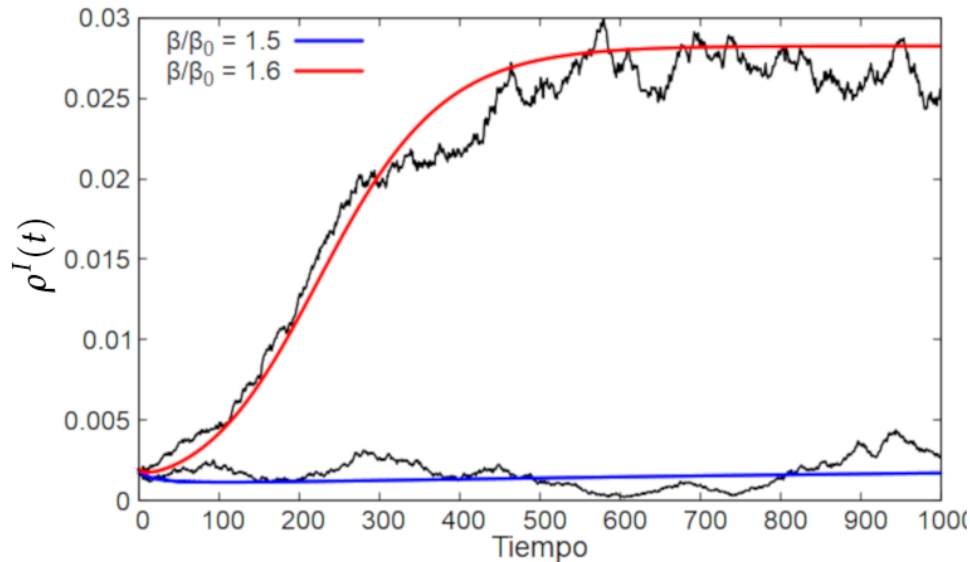


Figura 4.1: Simulaciones Monte Carlo (negro) frente a iteración de las ecuaciones markovianas (rojo y azul) para un modelo SIS, con distintos valores de probabilidad de contagio. Se representa el impacto de la epidemia  $\rho^I$  en función del tiempo. El valor de  $\beta$  está expresado en función de  $\beta_0$ , que es el umbral epidémico para un valor nulo de la movilidad,  $\beta_0 = \beta_c(p = 0)$ .

Vemos que la iteración de las ecuaciones markovianas lleva directamente al estado estacionario, ya que son deterministas, mientras que las simulaciones Monte Carlo oscilan en torno a ese estado. Nuestro objetivo es comprobar que el estado estacionario al que se llega coincide para ambos métodos, con distintos valores de  $\beta$  y  $p$ .

Algo importante que abordar es el coste computacional que conllevan estas simulaciones. Mientras que las iteraciones de las ecuaciones markovianas son rápidas porque hay  $N$  variables en total, las simulaciones Monte Carlo son mucho más demandantes. Tienen que simular el estado de probablemente millones de individuos y las interacciones entre ellos, por lo que esto es varios órdenes de magnitud más lento. Además, otro problema es que debido a la aleatoriedad de los estados hay que tomar muchos pasos temporales para calcular la media.

Este proceso es generalmente ergódico, por lo que podemos hacer la media de la cantidad de infectados sobre muchos pasos temporales, y la precisión aumentará con el tiempo. Sin embargo, cerca del umbral epidémico se rompe la ergodicidad ya que las fluctuaciones estocásticas pueden hacer que la epidemia se extinga de forma artificial. En ese caso, además del promedio temporal, debemos promediar sobre diferentes realizaciones de las simulaciones. Haciendo esto para distintos valores de  $\beta$  y  $p$  tenemos el diagrama de fase de la figura 4.2.

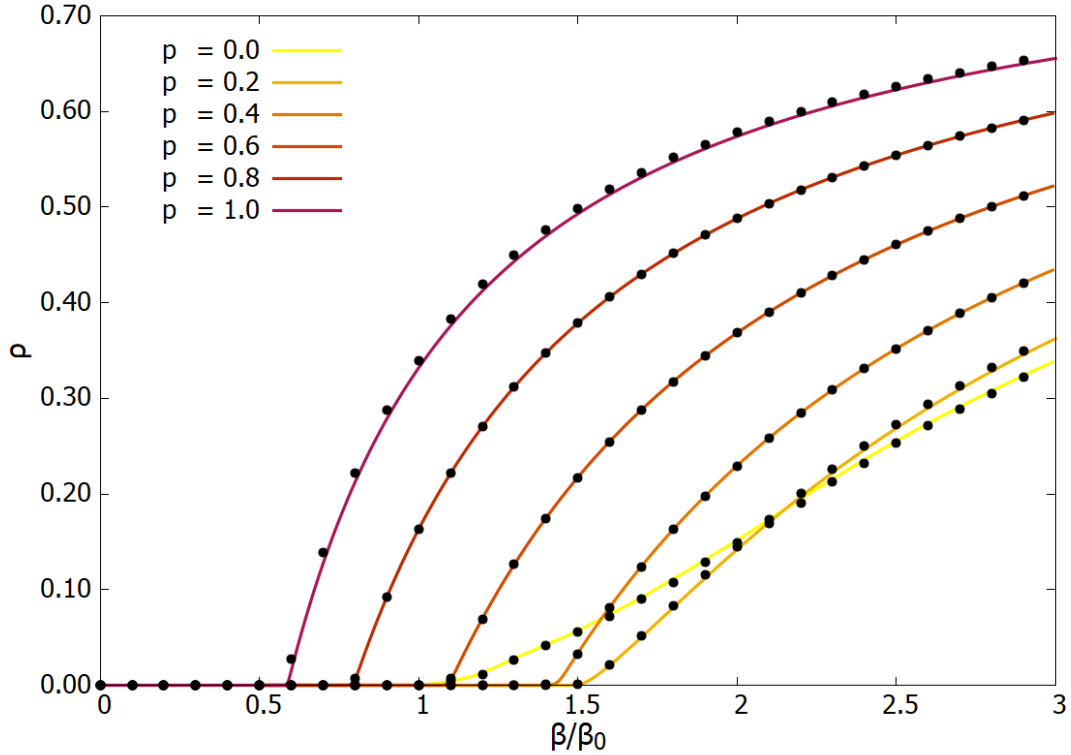


Figura 4.2: Diagrama de fases del modelo SIS para individuos indistinguibles, en la red de movilidad de Las Vegas. Las líneas continuas representan las iteraciones de las ecuaciones markovianas, y los puntos se corresponden a simulaciones Monte Carlo. Se representa el impacto de la epidemia en función de la probabilidad de contagio en el eje  $x$  y para distintos valores de movilidad  $p_d$ . Cada punto se ha obtenido calculando la media del impacto a lo largo de 300 pasos temporales (tras alcanzar el equilibrio), y 20 iteraciones.

En el eje  $x$  hemos representado la probabilidad de contagio normalizada a  $\beta_0$ , que es el umbral epidémico para el caso  $p = 0$ . Al no haber movilidad,  $\beta_0$  únicamente depende del nodo más poblado  $i$ . La expresión para el modelo SIS es simple:

$$\beta_0 = \frac{\mu}{n_i}. \quad (4.6)$$

Vemos cómo crece el impacto en función de la probabilidad de contagio, y cómo el valor de  $p$ , que representa la movilidad entre parches, juega un papel muy importante a la hora de estudiar la propagación de epidemias.

## 4.2. Cálculo del umbral epidémico

Una vez verificadas las ecuaciones, podemos ir adelante y obtener una expresión para el umbral epidémico, no únicamente para  $p = 0$  sino para cualquier valor del desplazamiento. Este desarrollo está ya explicado en [18], pero haremos un resumen. Básicamente, consiste en considerar que la cantidad de infectados es despreciable comparada con el total de la población y que nos encontramos en el estado estacionario. Esto se traduce en

$$\rho_i(t) \equiv \varepsilon_i(t) \ll 1, \quad (4.7)$$

$$\varepsilon_i(t) = \varepsilon_i^* \quad (4.8)$$

para todos los parches de la metapoblación. Usando esto, podemos linealizar las ecuaciones, despreciando todos los términos mayores que de primer orden, llegando a

$$\frac{\mu}{\beta} \varepsilon_i^* = (1-p)^2 n_j \varepsilon_i^* + \sum_{j=1}^N \left( p(1-p) n_j (R_{ij} + R_{ji}) + p^2 n_j \sum_{l=1}^N R_{il} R_{jl} \right) \varepsilon_j^*. \quad (4.9)$$

Esto se puede traducir en una ecuación de autovalores

$$\frac{\mu}{\beta} \varepsilon_i^* = (\mathbf{M} \bar{\varepsilon}^*)_i \quad (4.10)$$

donde la matriz es la siguiente:

$$M_{ij} = (1-p)^2 \delta_{ij} n_j + p(1-p) n_j (\mathbf{R} + \mathbf{R}^T)_{ij} + p^2 (\mathbf{R} \cdot \mathbf{R}^T)_{ij}. \quad (4.11)$$

Vemos que (4.10) es una ecuación de autovalores, donde el valor propio máximo de la matriz nos dará el valor mínimo de  $\beta$  posible para que la epidemia no se extinga, por lo que de esa forma obtenemos el umbral epidémico

$$\beta_c(p) = \frac{\mu}{\Lambda_{\max}(\mathbf{M}(p))}. \quad (4.12)$$

Gracias a las ecuaciones markovianas, hemos podido obtener una expresión analítica para el umbral en función de la probabilidad de desplazamiento. Es importante el hecho de que esta expresión se puede aplicar a la mayoría de enfermedades, porque depende principalmente de la forma de la red en concreto.

### 4.3. Problemas y limitaciones del formalismo

Este formalismo tiene un amplio margen de mejora, como se ha hecho ya de varias formas en [19], donde se introducen redes multicapa, en [20], añadiendo ciclos diurnos-nocturnos y restringiendo movilidad, o en [21], donde se trata de modelizar la propagación del COVID-19. Sin embargo, aún tiene problemas fundamentales que deben resolverse.

Para empezar, considerar a todos los nodos de una misma subpoblación como indistinguibles es poco realista. Las personas que residen en un área tienen comportamientos heterogéneos, con patrones de desplazamiento característicos y recurrentes. De esta forma, asumir que con cada paso temporal una persona decide aleatoriamente su destino de viaje no está justificado y debe ser corregido.

Por otro lado, la mayor parte de las fuentes de redes de movilidad están basadas en encuestas acerca de desplazamientos al trabajo, o a los lugares más frecuentados. Todo esto nos indica que podemos esperar que las tasas de infección dentro de un mismo parche sean diferentes en función del destino habitual de cada individuo, por lo que en el siguiente punto introducimos un formalismo que tendrá estas diferencias en cuenta.

## 5. Modelo de metapoblaciones con movilidad recurrente e individuos distinguibles

Este formalismo que llamaremos de individuos distinguibles será similar al anterior, pero en lugar de mover aleatoriamente a los agentes, asignamos lugares de destino fijos para cada individuo. De esta forma, todos los viajes serán recurrentes, y lo único que será producto de azar es la decisión de viajar o quedarse en su parche de residencia, dada por el parámetro de probabilidad de desplazamiento  $p$ .

Vemos que ahora no es suficiente representar el estado mesoscópico del sistema usando la cantidad de infectados en cada parche, porque incluso si dos individuos residen en el mismo lugar, al desplazarse continuamente a destinos distintos van a tener probabilidades distintas de infectarse. Por lo tanto, en lugar de usar el vector  $\rho_i^I(t)$ , tendremos que utilizar una matriz de tamaño  $N \times N$ , cuya componente  $\rho_{ij}^I(t)$  nos dará la proporción de agentes residentes del nodo  $i$  que tienen asignado el destino  $j$  que están infectados (en el estado I).

Vamos ahora a obtener las ecuaciones markovianas para la evolución temporal del sistema. En este formalismo, la matriz origen-destino que usaremos será  $n_{ij}$ , que representa la cantidad de personas residentes en el nodo  $i$  que tienen como destino asignado el  $j$ . Tener en cuenta que aquellas personas cuyo lugar más frecuentado esté en el mismo parche se incluirán en la componente  $n_{ii}$ , por lo que puede existir un enlace de un vértice consigo mismo.

La primera ecuación es similar al formalismo anterior:

$$\rho_{ij}^I(t+1) = (1 - \mu)\rho_{ij}^I(t) + (1 - \rho_{ij}^I(t))\Pi_{ij}(t). \quad (5.1)$$

Ahora  $\Pi$  será distinto dependiendo del destino asignado a cada individuo, porque es el único al que pueden viajar.

$$\Pi_{ij}(t) = (1 - p)P_i(t) + pP_j(t). \quad (5.2)$$

De nuevo, calculamos  $P_i(t)$  a partir de la cantidad de personas infectadas que viajan al nodo  $i$  con las que habrá un contacto, donde obtenemos la probabilidad complementaria a no infectarse con ninguno de los contactos que se dan con los demás agentes, y viene dado por

$$P_i(t) = 1 - \prod_{j=1}^N (1 - \beta \rho_{ji}^I(t))^{p n_{ji}} \cdot \prod_{j=1}^N (1 - \beta \rho_{ij}^I(t))^{p n_{ij}}. \quad (5.3)$$

Como puede verse, no son más complicadas que las ecuaciones de individuos indistinguibles. Sin embargo, veremos cómo al haber más variables el coste computacional será considerablemente mayor.

### 5.1. Verificación de las ecuaciones

Como hemos visto anteriormente, es importante comprobar que las ecuaciones markovianas coinciden con las simulaciones de agentes individuales, y dibujamos los diagramas de fases en las gráficas de la figura 5.1. Vemos que, efectivamente, los puntos (correspondientes a la media de simulaciones Monte Carlo) coinciden con las líneas continuas resultantes de iterar las ecuaciones markovianas, para distintos valores de  $\beta$  y  $p$ . El umbral epidémico para el caso  $p = 0$ , es decir  $\beta_0$  es el mismo que en el caso de individuos indistinguibles.

Vemos que el umbral epidémico, que es el punto en el que el impacto deja de ser cero, cambia dependiendo del valor de  $p$ . Por esa razón conviene obtener la expresión analítica y estudiar cómo depende de la movilidad y de la red, lo cual haremos en la siguiente parte. En concreto, un resultado que podemos observar de las gráficas es que el umbral epidémico es mayor que  $\beta_0$  para algunos valores de  $p$ , lo cual al principio puede no parecer intuitivo. Mientras que es de esperar que la movilidad aumente la facilidad con la que se propaga la epidemia, parece que puede ser detrimental en ciertos casos.

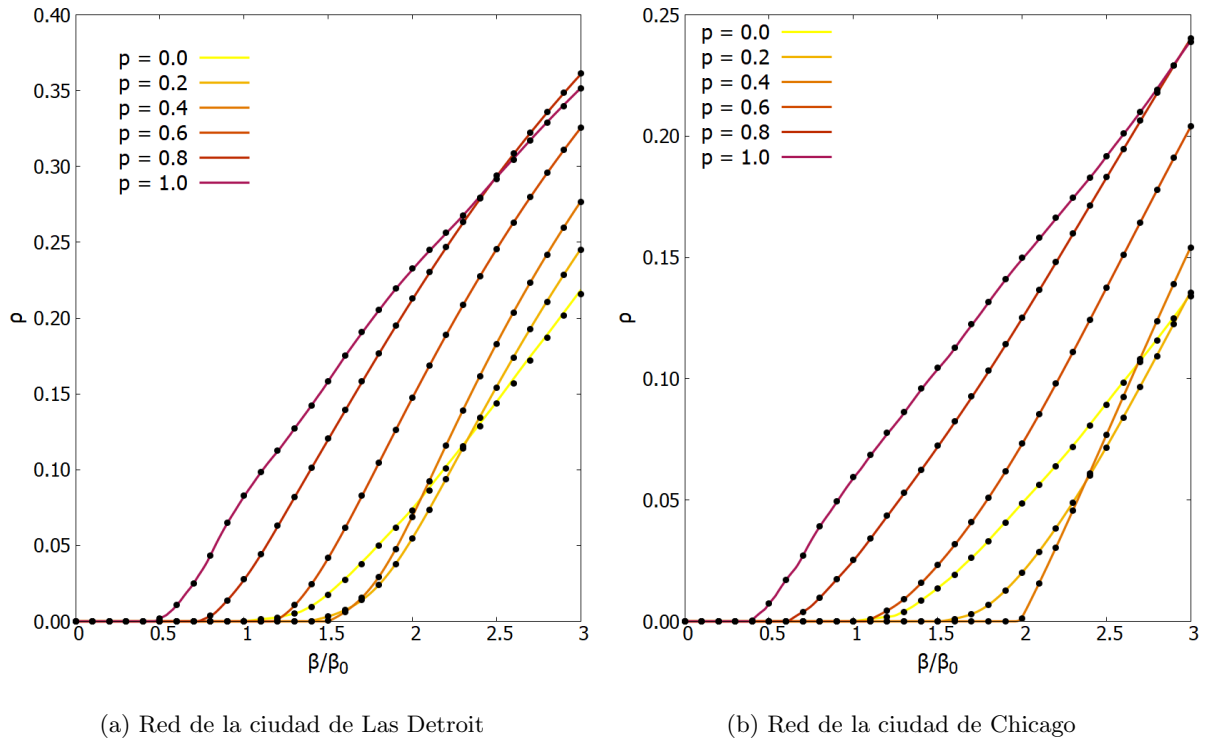


Figura 5.1: Diagrama de fases del modelo SIS para individuos distinguibles. Las líneas continuas representan las iteraciones de las ecuaciones markovianas, y los puntos se corresponden a simulaciones Monte Carlo.

## 5.2. Cálculo del umbral epidémico

Siguiendo la metodología vista en la sección 4.2, vamos a proceder a obtener una expresión analítica que nos proporcione el valor del umbral epidémico en función del valor de  $p$  y la matriz origen-destino. Primero, asumimos:

$$\rho_{ij}^I(t) \equiv \varepsilon_{ij}(t) \ll 1, \quad (5.4)$$

$$\varepsilon_{ij}(t) = \varepsilon_{ij}^*. \quad (5.5)$$

Ahora sustituimos en la ecuación (5.3) y eliminamos los términos de orden mayor que uno

$$P_i(\varepsilon^*) = 1 - \prod_{j=1}^N (1 - \beta \varepsilon_{ji}^*)^{p_d n_{ji}} \cdot \prod_{j=1}^N (1 - \beta \varepsilon_{ij}^*)^{(1-p_d) n_{ij}}$$

para después linealizar los dos productos por separado, y multiplicarlos más adelante.

$$\begin{aligned} \prod_{k=1}^N (1 - \beta \varepsilon_{ki}^*)^{p n_{ki}} &\sim \prod_{k=1}^N (1 - \beta p n_{ki} \varepsilon_{ki}^*) \sim 1 - \sum_{k=1}^N \beta p n_{ki} \varepsilon_{ki}^*, \\ \prod_{k=1}^N (1 - \beta \varepsilon_{ik}^*)^{p n_{ik}} &\sim \prod_{k=1}^N (1 - \beta p n_{ik} \varepsilon_{ik}^*) \sim 1 - \sum_{k=1}^N \beta p n_{ik} \varepsilon_{ik}^*, \\ P_i(\varepsilon^*) &= 1 - (1 - \sum_{k=1}^N \beta p \varepsilon_{ki}^* n_{ki}) \cdot (1 - \sum_{k=1}^N \beta (1-p) \varepsilon_{ik}^* n_{ik}) \sim \sum_{k=1}^N (\beta p \varepsilon_{ki}^* n_{ki} + \beta (1-p) \varepsilon_{ik}^* n_{ik}). \end{aligned}$$

Sustituyendo el valor de las  $P_i$  en (5.2) tenemos

$$\begin{aligned} \Pi_{ij} &= (1-p) \sum_{k=1}^N (\beta p \varepsilon_{ki}^* n_{ki} + \beta (1-p) \varepsilon_{ik}^* n_{ik}) + p \sum_{k=1}^N (\beta p \varepsilon_{kj}^* n_{kj} + \beta (1-p) \varepsilon_{jk}^* n_{jk}) = \\ &\quad \beta \sum_{k=1}^N \left( p(1-p) n_{ki} \varepsilon_{ki}^* + (1-p)^2 n_{ik} \varepsilon_{ik}^* + p^2 n_{kj} \varepsilon_{kj}^* + p(1-p) n_{jk} \varepsilon_{jk}^* \right). \end{aligned}$$

Por último, linealizamos la ecuación (5.1) teniendo en cuenta que  $\Pi_{ij}$  ya es de orden 1 con  $\varepsilon_{ij}^*$ , y sustituimos el resultado anterior en

$$\frac{\mu}{\beta} \varepsilon_{ij}^* = \sum_{k=1}^N (1-p)^2 n_{ik} \varepsilon_{ik}^* + p(1-p) n_{ki} \varepsilon_{ki}^* + p(1-p) n_{jk} \varepsilon_{jk}^* + p^2 n_{kj} \varepsilon_{kj}^*. \quad (5.6)$$

Esta es una ecuación de autovalores, que podemos reescribir de forma matricial como

$$\frac{\mu}{\beta} \varepsilon_{ij}^* = \sum_{k,l=1}^N M_{jk}^{il} \varepsilon_{lk}^* \quad (5.7)$$

donde, trabajando con los índices de los sumatorios, la matriz queda de la siguiente manera

$$M_{jk}^{il} = (1-p)^2 n_{ik} \delta_{il} + p(1-p) (n_{li} \delta_{ik} + n_{jk} \delta_{jl}) + p^2 n_{lj} \delta_{jk} \quad (5.8)$$

de forma que hemos hallado el umbral epidémico para el caso de individuos distinguibles, y viene dado por

$$\beta_c = \frac{\mu}{\Lambda_{max}(\mathbf{M})}. \quad (5.9)$$

A la hora de integrarla en un algoritmo para obtener los autovalores, la crearemos en dos dimensiones tal que  $M_{jk}^{il} \rightarrow M[i \cdot N + j][l \cdot N + k]$ . Los autovectores se corresponderán de la misma forma con  $\varepsilon_{ij} \rightarrow \varepsilon[i \cdot N + j]$

Esta matriz tiene mucha importancia, ya que representa los contactos que tiene una persona residente en el nodo  $i$  con destino  $j$ , con personas residentes en el nodo  $l$  y destino  $k$ . En el caso de individuos indistinguibles esta matriz tenía un tamaño  $N \times N$ , pero ahora es de tamaño  $N^2 \times N^2$ , por lo que ocupará mucha más memoria, y tendrá un coste computacional mayor para obtener los autovalores.

Sin embargo, no quiere decir que debamos almacenar  $N^4$  elementos. En el peor caso, la matriz tendrá  $4 \cdot N^3$  valores distintos de cero. Además, hay que recordar que la mayoría de redes con las que trabajaremos son esparcidas, y esto hará que solamente  $4 \cdot N \cdot L$  elementos sean distintos de cero.

Aún podemos reducir más el número de elementos. Cuando un enlace de la matriz origen-destino no existe, es decir  $n_{ij} = 0$ , el porcentaje de infectados de ese grupo no afecta al resto (porque al no haber personas, no pueden infectar a otros). Esto se traduce en que la columna  $i \cdot N + j$  de la matriz  $\mathbf{M}$  es nula, e implica que si forzamos la fila  $i \cdot N + j$  a que también se anule, los autovalores no cambian. Lo único que cambiarán son los autovectores, cuya componente  $i \cdot N + j$  se anulará. Como esas componentes se corresponden a enlaces no existentes de la matriz origen-destino, no tendremos interés en mantenerlas.

De esta forma, podemos llegar a reducir finalmente el número de elementos distintos de cero de la matriz a  $4 \cdot L^2/N$  en el mejor caso, que es mucho menos que el original de  $N^3$  si la matriz origen-destino es muy dispersa. Si pensamos en grafos regulares de grado  $k$ , que tienen un número constante de enlaces por vértice, la matriz escalaría con  $4 \cdot k^2 \cdot N$ . Tener en cuenta que la mayoría de redes van a ser algo más densas, y terminarán ocupando mucho espacio aún con todo esto.

Vamos ahora a comprobar que hemos obtenido la expresión correcta del umbral epidémico. Para ello, creamos un mapa de calor, representando el impacto de la epidemia en función de  $\beta$  y  $p$ , iterando las ecuaciones markovianas hasta alcanzar el estado estacionario. Podemos compararlo con la expresión del umbral epidémico que acabamos de obtener, en función de  $p$ . Mostramos el resultado en la figura 5.2.

En efecto, el umbral epidémico marca la separación entre las dos fases, siendo la zona en negro donde no puede propagarse la epidemia y la naranja en la que los contagios pueden alcanzar un equilibrio endémico. Observamos que, al aumentar el valor de  $p$ , ocurre un **detrimento** inicial que impide que la epidemia exista incluso para valores de  $\beta$  del doble que la inicial.

Se puede demostrar que este resultado es universal, independientemente de la forma de la red con la que trabajemos. A primer orden, el umbral epidémico siempre va a aumentar con  $p$  de la misma forma. La explicación tiene que ver con el hecho de que en el umbral epidémico en  $p = 0$ , la epidemia está completamente focalizada en un parche de la metapoblación. Al aumentar  $p$ , ocurren dos cosas importantes: Primero, residentes del foco, los principales infectados, viajan a otros parches menos poblados, donde la enfermedad no puede propagarse. Por lo tanto, todo el tiempo que no estén en el foco, conlleva posibles contagios que dejan de ocurrir, y supone un detrimento a la epidemia. Segundo, individuos de otros nodos se mueven hacia el foco, posiblemente contrayendo la enfermedad, y aumentando el número de contagios. Sin embargo, esta contribución es despreciable respecto a la primera, porque una vez la han contraído, es muy improbable que vuelvan al foco (recordar  $p$  es cercano a cero), por lo que se quedan en sus parches de residencia y no contribuirán a propagar la epidemia.

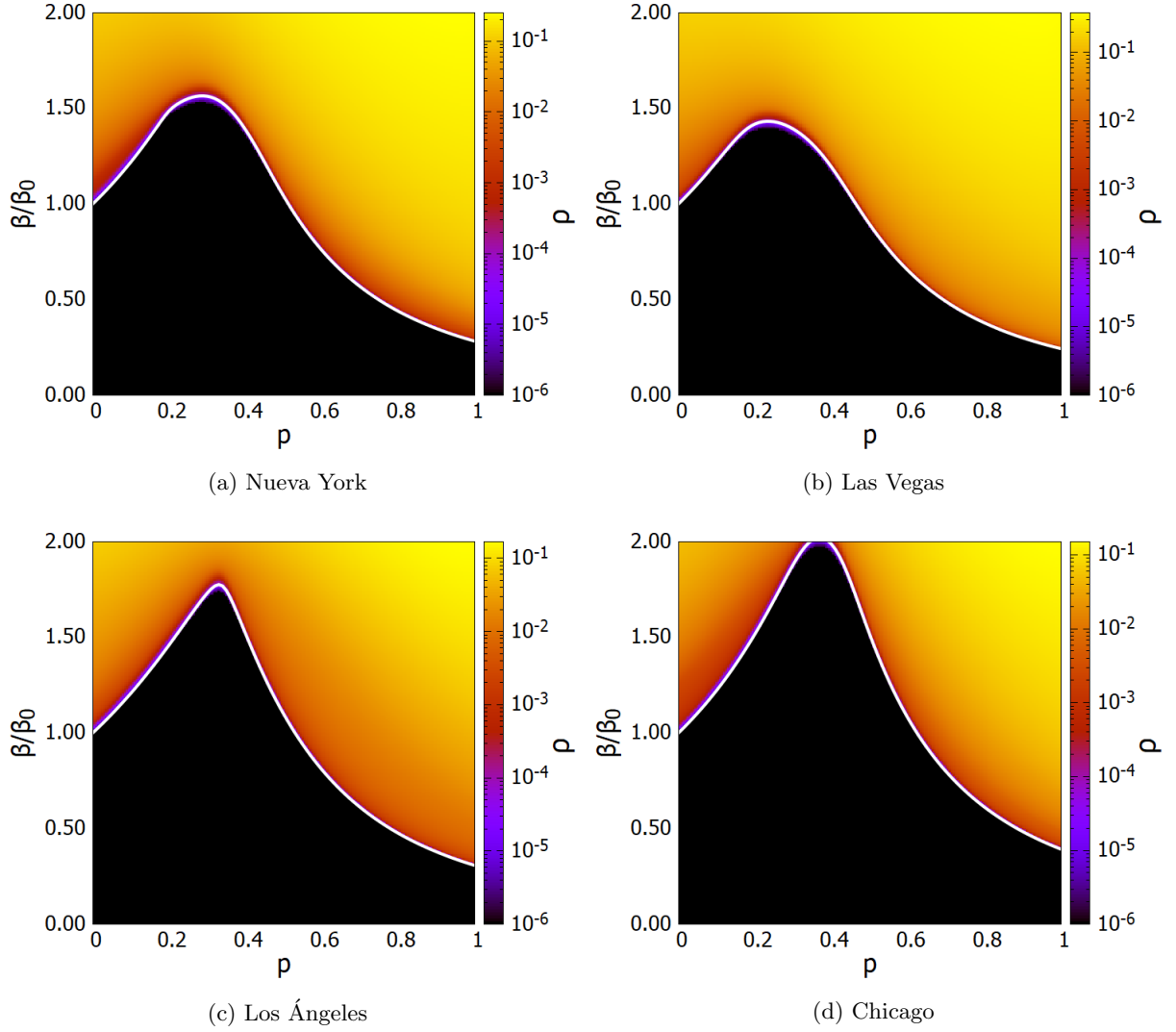


Figura 5.2: Mapas de calor del impacto de la epidemia en función de la movilidad (eje horizontal) y la probabilidad de contagio (eje vertical), a partir de iterar las ecuaciones markovianas. La línea blanca representa el umbral epidémico  $\beta_c(p)$ .

Un patrón recurrente que vamos a observar a lo largo del trabajo es la relevancia de la heterogeneidad de los sistemas. Cuando tratamos con epidemias, una red muy heterogénea va a implicar tener umbrales epidémicos más bajos, y una mayor facilidad para que exista una epidemia incluso para  $\beta$  baja. En la práctica, esta es realmente la mayor diferencia entre los formalismos indistinguible y distinguible. En el primero, la aleatoriedad de los viajes de cada individuo va a forzar una homogeneización de la población que va a llevar a umbrales epidémicos mayores, incluso si la cantidad de personas que viajan entre nodos es exactamente la misma en ambos. Esto remarca la importancia del hecho de que los desplazamientos de las personas son recurrentes.

A continuación vamos a ver cómo aplicar lo que hemos visto a políticas de control de epidemias.



## 6. Flujos críticos de movilidad humana

Una vez reconozcamos la importancia de la movilidad humana como factor imprescindible en el estudio de propagación de enfermedades, mostraremos cómo uno de los resultados anteriores puede ayudarnos a llevar a cabo políticas de movilidad y control epidémico. Empezaremos hablando del vector propio asociado al mayor autovalor de la matriz de mezcla  $\mathbf{M}$ .

Este vector representa las proporciones de infectados de todos los enlaces, que están asociados a los conjuntos de personas que viajan de un nodo  $i$  a otro parche  $j$ . Es de esperar entonces que nos dé información muy valiosa de los flujos más relevantes y que tienen un mayor impacto en el contagio de la enfermedad.

Partiendo de que los datos de movilidad están basados principalmente en viajes y desplazamientos diarios, y tienen en cuenta las personas que viajan a su propio parche, consideraremos que el valor de  $p$  del que partimos es cercano a 1. Durante esta sección, usaremos  $p = 0,7$ . La primera forma de aumentar el umbral epidémico es simplemente reducir el valor de  $p$ , hasta el punto en el que el umbral alcanza su máximo, como vemos en las figuras en Figure 5.2. Esto supone reducir los viajes indiscriminadamente de todos los agentes en la metapoblación.

Mientras que esta es una respuesta razonable, cabe preguntarse si podemos seleccionar los flujos más importantes y restringir su movilidad, para aumentar el umbral epidémico desplazando a un número mucho menor de personas.

Vamos a escoger cuatro criterios a la hora de elegir los enlaces que vamos a restringir:

- Aleatorio: para tener modelo nulo con el que comparar el resultado, creamos un algoritmo que selecciona aleatoriamente los enlaces que van a ser eliminados. Esto es similar, aunque no equivalente, a reducir el valor de  $p$ .
- Población: la epidemia va a encontrarse en los focos que acumulan a un número mayor de personas. Por lo tanto, podemos esperar buenos resultados si nos centramos en cortar los flujos que mueven a más agentes primero.
- Mayores componentes del vector propio: usaremos las propiedades espectrales de la matriz para determinar los flujos con mayor porcentaje de infectados.
- Mayor cantidad de infectados: de nuevo, usando el vector propio, pero lo multiplicamos por la cantidad de personas asociadas a ese enlace, para centrarnos en los flujos que mueven a más individuos.

La forma en la que llevamos a cabo el algoritmo será la siguiente: primero, resolvemos la ecuación (4.10) y obtenemos el vector propio asociado al autovalor mayor. Después, seleccionamos el enlace que queramos modificar siguiendo uno de los criterios anteriores, y modificamos la matriz origen-destino de forma que reducimos el valor de  $n_{ij}$  en una cantidad que prefijemos (por ejemplo, 2000 personas). Añadimos esa cantidad al valor de  $n_{ii}$ , ya que no estamos confinando a estas personas, sino restringiendo su movilidad a otros nodos, pero siguen siendo parte del sistema. Si no hay suficientes personas asociadas a ese enlace, se busca otro siguiendo el criterio correspondiente, y sucesivamente.

Una vez hemos restringido el movimiento de las 2000 personas nuestra red es distinta a la original, y tanto el umbral epidémico como los flujos críticos habrán cambiado. Por ello, volvemos a resolver la ecuación de autovalores y obtenemos otro vector propio. Repetimos el proceso para grupos de 2000 personas, registrando la evolución del umbral cada vez que resolvemos el sistema. Vemos los resultados en la figura 6.1.

En efecto, todos los métodos tienden a aumentar el umbral epidémico, lo cual es de esperar ya que partimos de un umbral menor que  $\beta_0$  y eliminar todos los enlaces de la red es equivalente a estar en el caso  $p = 0$ . Los tres criterios mencionados son considerablemente mejores que eliminar enlaces aleatoriamente, pero los que están basados en las componentes del vector propio dan resultados excelentes en comparación al que se centra simplemente en la población. No solo hacen que aumente considerablemente el umbral epidémico restringiendo a una porción pequeña de la población, sino que además superan en muchos casos el valor de  $\beta_0$ .

En ocasiones, cortar un enlace supone una disminución del umbral epidémico, y este hecho nos da una pista de lo que ocurre a nivel interno en la red. La enfermedad suele estar concentrada en un foco principal, que atrae más agentes que el resto. Una vez cortamos los enlaces que viajan a éste, la epidemia se descentraliza y un segundo parche se convertirá en el nuevo foco. En las gráficas, este momento se puede observar en el punto donde las dos curvas correspondientes a los criterios del vector propio se separan, porque hasta entonces el efecto de eliminar cualquier enlace que viaje al foco es el mismo.

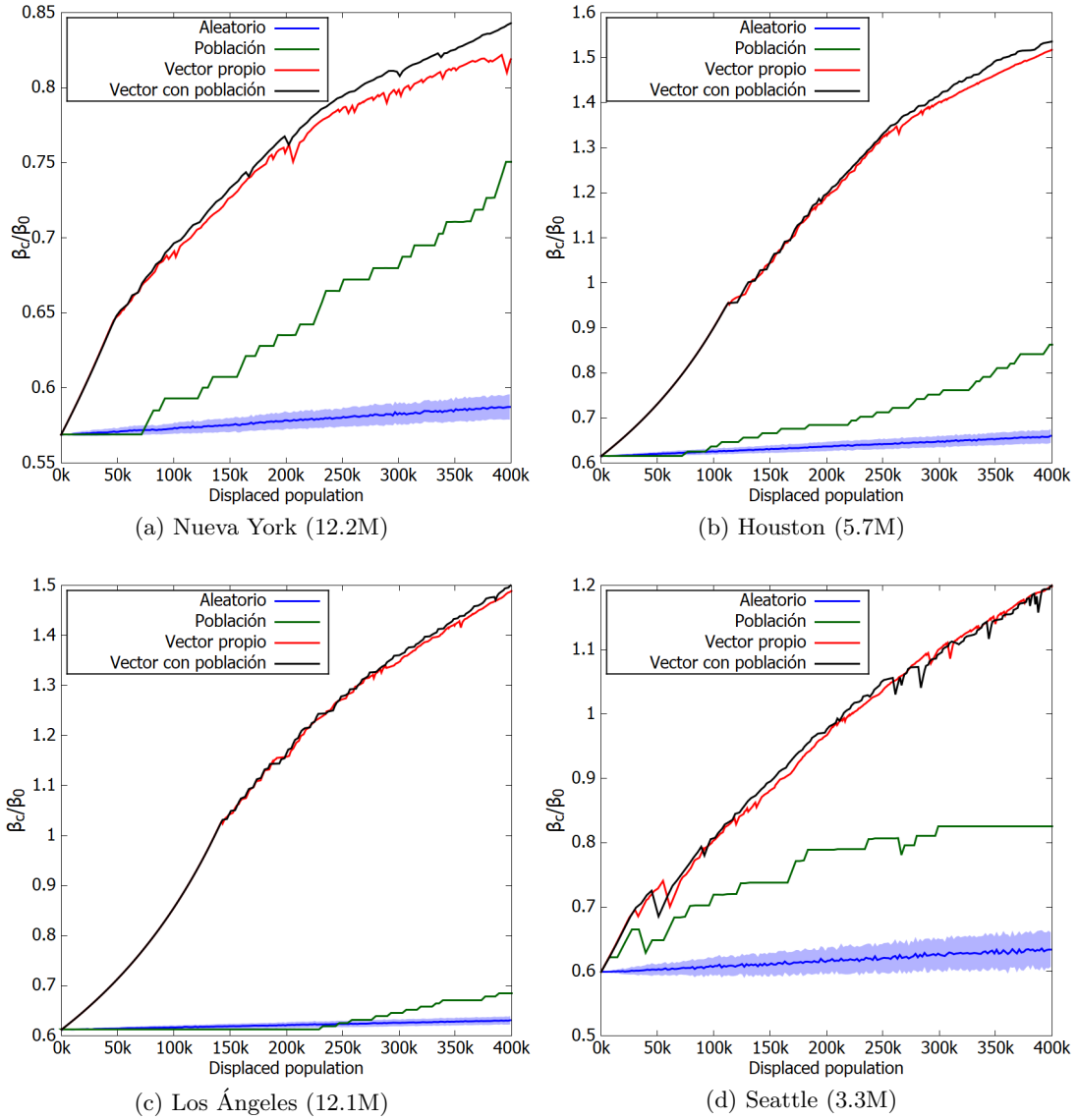


Figura 6.1: Umbral epidémico en función de la cantidad de personas a las que se ha restringido la movilidad, siguiendo los criterios explicados. Del algoritmo asociado al criterio aleatorio, se representa la media y la desviación que abarca el 95 % de las muestras.

En el caso de que ambos estén conectados, cortar su enlace común moverá todas las personas de uno de ellos al segundo, que aumentará considerablemente su población efectiva y sus contactos, reduciendo de esta forma el valor del umbral. Si seguimos cortando enlaces, el número de parches que empiezan a competir por ser el foco principal aumenta junto con el umbral epidémico.

Vemos de nuevo la importancia de la heterogeneidad en la red. El umbral se encuentra en su punto más bajo cuando un único parche es responsable de la mayor parte de los contagios, y es mayor cuando el foco está repartido entre diversos nodos.

Otro aspecto a comentar es el número de personas que desplazamos antes de volver a calcular los valores y vectores propios, que hemos definido como 2000 previamente. Idealmente, cuanto menor es esta cantidad mejor va a ser el resultado, pero hay que tener en cuenta el coste computacional asociado a resolver una matriz de este tamaño. Para redes como la de Nueva York, que consta de 147 mil enlaces, el tiempo de cada paso es del orden de minutos, lo cual es una limitación importante.

Vistas las posibilidades que nos presenta el nuevo formalismo, nos centraremos en desarrollar a partir del mismo un modelo más elaborado que nos permita abordar casos realistas, centrándonos en la capacidad de replicar enfermedades tales como el COVID-19. En la sección a continuación veremos cómo implementar un modelo compartimental más complejo, y mejorar partes de los formalismos tanto distinguible como indistinguible para adaptarse mejor a los patrones de movilidad entre personas y los contactos que ocurren entre ellos.

## 7. Desarrollando un modelo más complejo

Hasta ahora, nos hemos basado en el modelo compartimental SIS y en que los agentes tienen contactos con todos los demás individuos dentro de un nodo. Esto, a pesar de permitirnos estudiar el impacto de una epidemia en función de la forma de las distintas redes y la importancia de la movilidad y los distintos flujos de personas, no es suficientemente elaborado como para poder replicar un caso real. Para ello, tendremos que llevar a cabo ciertas mejoras que sean, en la medida de lo posible, fáciles de generalizar. Vamos a empezar por la forma en la que medimos los contactos en una subpoblación.

### 7.1. Densidad de población y función de contactos

Aunque es cierto que parches más poblados, en general, van a ver una cantidad mayor de contactos entre agentes, esta es una simplificación poco realista, especialmente teniendo en cuenta la variabilidad de la escala de los datos que tenemos a nuestra disposición. El primer cambio que haremos será empezar a considerar los contactos como función de la densidad de una población, en lugar de la cantidad de personas.

Por simplicidad, vamos a considerar que la cantidad de contactos que una persona tiene con agentes en la subpoblación en la que se encuentre en el momento de la reacción es proporcional a la densidad de población en dicho lugar como

$$f\left(\frac{n_i^{eff}}{a_i}\right) = \frac{n_i^{eff}}{a_i} \quad (7.1)$$

siendo  $a_i$  el área de la subpoblación  $i$ , y  $n_i^{eff}$  la población efectiva que se encuentra en ella, que definiremos más adelante. Notar que hemos usado un signo de igualdad porque añadiremos a continuación el factor de normalización, para controlar la cantidad de contactos totales que queremos que se lleven a cabo.

Uno de nuestros objetivos a tener en cuenta cuando desarrollemos el modelo será permitir que esta función sea completamente generalizable, y que las expresiones que derivemos más adelante no dependan de cómo la definamos. Existen estudios como [22] en los que se estudia cómo cambia el número de contactos entre individuos con las densidades de población, pero no entraremos en ello en este trabajo. Basta con asegurarnos de que en cualquier momento podamos sustituir la función por otra más adecuada de forma inmediata.

Para fijar el número de contactos que se dan en una población utilizábamos anteriormente la cantidad total de personas en el parche, que podían llegar a ser miles, lo cual es una cantidad demasiado alta. En general, el número real ronda los 8 contactos en el entorno de trabajo y 3 en el lugar de residencia [23].

Para controlar estos números, vamos a introducir el parámetro de normalización  $z$  a partir de un  $\langle k \rangle$  que elijamos, tal que

$$z = \frac{\sum_{i=0}^N n_i^{eff} \langle k \rangle}{\sum_{i=0}^N n_i^{eff} f\left(\frac{n_i^{eff}}{a_i}\right)} \quad (7.2)$$

A partir de ahora, para simplificar la notación, usaremos  $f_i = f(\frac{n_i^{eff}}{a_i})$ . De esta forma, si el número de contactos es

$$k_i = z \cdot f_i \quad (7.3)$$

la media de contactos por individuo en la totalidad de la población será  $\langle k \rangle$ , que podremos escoger sin importar la forma de la red o la función  $f$ .

## 7.2. Procesos de reacción diurnos y nocturnos

Al contrario que el formalismo de individuos indistinguibles, el de agentes distinguibles tiene un problema muy serio que pone en compromiso su capacidad de modelar un sistema realista para  $p$  cercana a 1. Ocurre que, tal y como hemos definido el formalismo, solamente ocurre el proceso de reacción tras haber desplazado los agentes a sus correspondientes localizaciones de destino. Esto significa que, para  $p = 1$ , los individuos siempre se encontrarán en el mismo parche (su destino asignado) en el momento de que ocurran los contagios.

Entonces, este caso no es distinto a una red en la que consideramos el destino de estos agentes como su residencia y nos ponemos en el caso  $p = 0$ . No existe movilidad entre parches, no importa la estructura de la red, simplemente es un modelo de poblaciones estáticas y que no interactúan entre ellas. De hecho, ni siquiera tendrán contactos los agentes que residen en el mismo parche. No podemos caracterizar de esta forma sistemas reales donde consideramos  $p = 1$  el punto en el que mayor viajes existen entre subpoblaciones.

Notar que este problema no es tan grave en el formalismo de individuos indistinguibles, en el que en cada paso temporal un agente elegirá un destino diferente, y además tendrá oportunidad de interactuar con todos los individuos de su misma subpoblación de residencia. A pesar de esto, también aplicaremos estos cambios a este formalismo para poder comparar ambos más adelante.

La forma en la que resolveremos este problema será separar el proceso de reacción en dos fases. La primera será exactamente igual a lo que hacíamos anteriormente: desplazar a las personas a sus destinos y dejar que ocurran los contagios. Después, devolvemos a los agentes a sus lugares de residencia, y allí volvemos a tener un proceso de contagio, de forma equivalente a un caso  $p = 0$ . Este mecanismo replica la dinámica de las personas que por el día se desplazan a sus destinos (trabajo, universidad, ocio) y por la noche vuelven a sus residencias, contactando con familiares o vecinos.

Aunque el proceso de contagio en ambos casos se da de la misma forma, asignaremos un número de contactos medio diferente a cada uno. Usaremos el superíndice  $D$  para denotar los parámetros correspondientes a las reacciones diurnas y  $N$  para las nocturnas:

$$\langle k^D \rangle = 8, \quad \langle k^N \rangle = 3. \quad (7.4)$$

Asumimos que persona que se hay infectado por el día no puede infectar por la noche. Esto no nos permitiría usar una cadena de Markov con un paso diario, necesitaríamos dos. Esto no supone una pérdida de realismo, ya que las personas no son contagiosas desde el momento en el que se exponen a la enfermedad. En el modelo compartimental que usaremos hay un estado intermedio en el que la persona expuesta no puede infectar a otros. De la misma forma, no hay dos procesos de recuperación con probabilidad  $\mu$  por el día y noche, sino que sigue habiendo solamente uno tras los contagios nocturnos.

Algo más que añadir es que no utilizaremos la función de densidad explicada anteriormente para calcular los contactos nocturnos, sino que usaremos un parámetro  $\sigma_i$ , que es la cantidad media de personas residentes en el hogar para un parche  $i$ . Sirve para generalizar el modelo en caso de que existan datos concretos sobre el tamaño de los hogares en cada parche. Los contactos diurnos y nocturnos quedan como:

$$k_i^D = z^D \cdot f_i, \quad z^D = \frac{\sum_{i=0}^N n_i^{eff} \langle k^D \rangle}{\sum_{i=0}^N n_i^{eff} f_i} \quad (7.5)$$

$$k_i^N = z^N \cdot \sigma_i, \quad z^N = \frac{\sum_{i=0}^N n_i \langle k^N \rangle}{\sum_{i=0}^N n_i \sigma_i} \quad (7.6)$$

En ausencia de datos disponibles, asumiremos que el tamaño de los hogares  $\sigma$  es igual para todos los parches, por lo que  $z^N = \langle k^N \rangle / \sigma$  y por lo tanto  $k_i^N = \langle k^N \rangle$  para todo  $i$ .

### 7.3. Modelo compartimental SEAPIDR

Hemos presentado este modelo anteriormente en 3.3.3, y es el que usaremos a partir de ahora, ya que es apropiado para estudiar enfermedades como el COVID-19. Por supuesto, estos modelos pueden complicarse indefinidamente, añadiendo variables (distintos efectos en personas dependiendo del sexo y grupo de edad, nuevos compartimentos, vacunas, pérdida de inmunidad, ocupación de UCI, etcétera) o complejidad (competencia entre cepas, medidas de confinamiento), pero están fuera del alcance de este trabajo.

Vamos a ver cómo implementamos este modelo en un sistema de un único parche, para obtener una expresión del valor de  $\beta_0$ , que es el umbral epidémico para  $p = 0$ . Para el modelo SIS esto era inmediato, pero aquí requiere un poco más de trabajo. Es importante tener una expresión simple de este parámetro porque, al realizar las simulaciones y las comprobaciones más adelante, el rango más interesante que estudiar es el que se encuentra en la proximidad del umbral ( $0 \leq \beta \leq 3\beta_0$ ). Sin embargo, calcularlo a partir de la expresión general del umbral epidémico es muy costoso. Además, en  $p = 0$  los formalismos distinguible e indistinguible son equivalentes, por lo que este resultado nos servirá para ambos.

Para calcular el umbral epidémico, tanto ahora como cuando lo hagamos con cualquier  $p$ , seguiremos un procedimiento más sistemático descrito en [24], que se centrará en la obtención del **número básico de reproducción**  $R_0$ . En esta sección explicaremos los conceptos básicos acerca de esta magnitud y cómo obtenerla analíticamente.

#### 7.3.1. El número básico de reproducción $R_0$

$R_0$  se define como el número promedio de casos nuevos que genera un caso dado a lo largo de un período infeccioso al comienzo de una epidemia, es decir, cuando la totalidad de la población es susceptible. De esta forma, si  $R_0 > 1$  significa que cada infectado podrá contagiar a más de una persona antes de recuperarse, por lo que la enfermedad se propagará por la población de forma exponencial. Si, por el contrario,  $R_0 < 1$  se extinguirá. Con esta definición, en el momento en que la proporción de susceptibles en una población decrezca por debajo de  $1/R_0$  la epidemia no podrá propagarse.

¿Qué relación tiene  $R_0$  con el umbral epidémico y la probabilidad de contagio  $\beta$ ? Ambas magnitudes representan un concepto casi idéntico, pero lo hacen teniendo en cuenta distintos factores.

Mientras que la probabilidad de contagio es una magnitud intrínseca de la enfermedad, que no varía dependiendo de otros factores, el valor de  $R_0$  solamente tiene sentido cuando tenemos en cuenta las circunstancias y elementos externos. Es distinta dependiendo de la topología de la red, la cultura, las medidas tomadas por los gobiernos, etcétera. Por eso, a pesar de ser la misma enfermedad, el COVID-19 se ha propagado de forma diferente dependiendo de su localización. Hay muchos factores que pueden afectar al valor de  $R_0$  e incluso llevarlo a ser menor que 1.

Esto significa que, para un mismo sistema, ambas definiciones van a ser equivalentes. Con el procedimiento que seguiremos a continuación, seremos capaces de obtener el número básico de reproducción  $R_0$  en función de la probabilidad de contagio  $\beta$ , y entonces despejaremos  $\beta(R_0)$  siempre que sea posible (veremos que lo es, porque en este caso  $\beta$  y  $R_0$  son directamente proporcionales). Hallaremos el umbral epidémico definiéndolo como la probabilidad de contagio para la cual el valor de  $R_0$  de la epidemia se vuelve igual a 1, es decir  $\beta_c = \beta(R_0 = 1)$ . Vamos a ver cómo calculamos el número básico de reproducción.

### 7.3.2. Calculando $R_0$ para el modelo compartimental SEAPIDR

El primer paso consiste en partir de las ecuaciones no lineales de la evolución del sistema, y linealizarlas. Son las mismas que hemos presentado en el apartado 3.3.3, pero las escribimos de nuevo por conveniencia:

$$\begin{aligned}\rho^E(t+1) &= (1 - \eta)\rho^E(t) + (1 - \rho^E(t) - \rho^A(t) - \rho^P(t) - \rho^I(t) - \rho^D(t) - \rho^R(t))\Pi(t), \\ \rho^A(t+1) &= (1 - \mu_A)\rho^A(t) + (1 - x)\eta\rho^E(t), \\ \rho^P(t+1) &= (1 - \alpha)\rho^P(t) + x\eta\rho^E(t), \\ \rho^I(t+1) &= (1 - \delta - \mu_I)\rho^I(t) + \alpha\rho^P(t), \\ \rho^D(t+1) &= (1 - \gamma)\rho^D(t) + \delta\rho^I(t), \\ \rho^R(t+1) &= \mu_A\rho^A(t) + \mu_I\rho^I(t) + \gamma\rho^D(t), \\ \Pi(t) &= 1 - \left(1 - \beta_A\rho^A(t) - \beta_P\rho^P(t) - \beta_I\rho^I(t)\right)^k.\end{aligned}\tag{7.7}$$

Antes que nada, tener en cuenta que el factor  $k$  es extremadamente importante, ya que define la cantidad de contactos que cada individuo tiene en el campo medio del nodo en el que se encuentra. En el caso de un único parche, será  $k = \langle k^D \rangle + \langle k^N \rangle$ .

Ver que todas las ecuaciones son ya lineales, excepto la primera y la última. Asumiendo  $\rho^E(t) + \rho^A(t) + \rho^P(t) + \rho^I(t) + \rho^D(t) + \rho^R(t) \ll 1$  podemos aproximar

$$\Pi(t) \approx k \cdot \left(\rho^A(t) + \rho^P(t) + \rho^I(t)\right)\tag{7.8}$$

y sustituirlo en la primera ecuación, tal que

$$\rho^E(t+1) \approx (1 - \eta)\rho^E(t) + k \cdot \left(\rho^A(t) + \rho^P(t) + \rho^I(t)\right).\tag{7.9}$$

A partir de aquí, hay que construir lo que se conoce como la *next-generation matrix* o NGM, que llamaremos  $\mathbf{K}$ , la cual relaciona las cantidades de individuos que se encuentran en distintos compartimentos. Se puede descomponer en  $\mathbf{T} + \mathbf{\Sigma}$ , donde  $\mathbf{T}$  es la parte que describe la **transmisión**, es decir, el contagio de nuevos agentes, y  $\mathbf{\Sigma}$  es la parte de **transición**, que explica los cambios de los individuos ya infectados de un compartimento a otro. Es claro entonces que la primera ecuación de (7.7) nos dará la matriz de transmisión y las demás serán la de transición.

Otro aspecto a tener en cuenta es que los últimos compartimentos pueden omitirse, ya que no influyen en el proceso de contagio, simplemente tienen unos ritmos constantes de cambio en función de los compartimentos que vienen antes que ellos. Así, nos quedamos únicamente con  $s \in \{E, A, P, I\}$ , en cuyo orden definimos el vector de estado del sistema. Debemos escribir ahora las ecuaciones que describen el **cambio** de los infectados en cada compartimento, es decir,  $\Delta\rho^s(t+1) \equiv \rho^s(t+1) - \rho^s(t)$ .

$$\begin{aligned}\Delta\rho^E(t+1) &= -\eta\rho^E(t) + k \cdot (\rho^A(t) + \rho^P(t) + \rho^I(t)), \\ \Delta\rho^A(t+1) &= -\mu_A\rho^A(t) + (1-x)\eta\rho^E(t), \\ \Delta\rho^P(t+1) &= -\alpha\rho^P(t) + x\eta\rho^E(t), \\ \Delta\rho^I(t+1) &= -(\delta + \mu_I)\rho^I(t) + \alpha\rho^P(t).\end{aligned}\tag{7.10}$$

Podemos ahora escribir las matrices. La matriz  $\mathbf{T}$  viene dada por la primera ecuación de (7.10) excepto por el término de transición  $-\eta\rho^E(t)$ . El elemento  $i, j$  viene dado por la cantidad nuevos infectados en el compartimento  $i$  debido a contagios aportados por el compartimento  $j$  tal que

$$\mathbf{T} = k \cdot \begin{pmatrix} 0 & \beta_A & \beta_P & \beta_I \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.\tag{7.11}$$

De la misma forma, la matriz de transición se construirá con el resto de términos, donde el elemento  $i, j$  es el cambio de infectados en el compartimento  $i$  en función de los que hay en el  $j$ . Como todos ellos se vacían automáticamente, la diagonal estará formada por elementos negativos que representan el ratio con el que los individuos salen de cada estado.

$$\Sigma = \begin{pmatrix} -\eta & 0 & 0 & 0 \\ (1-x)\eta & -\mu_A & 0 & 0 \\ x\eta & 0 & -\alpha & 0 \\ 0 & 0 & \alpha & -(\delta + \mu_I) \end{pmatrix}.\tag{7.12}$$

Pues bien, una vez tenemos las dos matrices, el método afirma que el valor de  $R_0$  es el máximo autovalor de la matriz  $\mathbf{K}_L = -\mathbf{T}\Sigma^{-1}$ . Calculamos la inversa de la matriz  $\Sigma$ , que es

$$\Sigma^{-1} = \begin{pmatrix} \frac{1}{\eta} & 0 & 0 & 0 \\ \frac{1-x}{\mu_A} & \frac{1}{\mu_A} & 0 & 0 \\ \frac{x}{\alpha} & 0 & \frac{1}{\alpha} & 0 \\ \frac{x}{\delta + \mu_I} & 0 & \frac{1}{\delta + \mu_I} & \frac{1}{\delta + \mu_I} \end{pmatrix}\tag{7.13}$$

y multiplicándolas podemos calcular  $\mathbf{K}_L$  como



$$\begin{aligned}
\mathbf{K}_L = -\mathbf{T}\Sigma^{-1} = k \cdot & \begin{pmatrix} 0 & \beta_A & \beta_P & \beta_I \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{\eta} & 0 & 0 & 0 \\ \frac{1-x}{\mu_A} & \frac{1}{\mu_A} & 0 & 0 \\ \frac{x}{\alpha} & 0 & \frac{1}{\alpha} & 0 \\ \frac{x}{\delta + \mu_I} & 0 & \frac{1}{\delta + \mu_I} & \frac{1}{\delta + \mu_I} \end{pmatrix} = \\
& k \cdot \begin{pmatrix} \beta_A \frac{1-x}{\mu_A} + \beta_P \frac{x}{\alpha} + \beta_I \frac{x}{\delta + \mu_I} & \frac{\beta_P}{\mu_A} & \frac{\beta_P}{\alpha} + \frac{\beta_I}{\delta + \mu_I} & \frac{\beta_I}{\delta + \mu_I} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (7.14)
\end{aligned}$$

por lo que el máximo autovalor es

$$R_0 = k \cdot \left( \beta_A \frac{1-x}{\mu_A} + \beta_P \frac{x}{\alpha} + \beta_I \frac{x}{\delta + \mu_I} \right). \quad (7.15)$$

Al tener tres probabilidades de contagio diferentes, tenemos dos grados de libertad para dar con un umbral epidémico. Por ello, asumiremos que están relacionadas entre sí de la siguiente forma

$$\beta \equiv \beta_P = \beta_I = 2\beta_A \quad (7.16)$$

y con esto podemos despejar un valor de  $\beta$  y para hallar el umbral epidémico nos situamos en el caso  $R_0 = 1$ .

$$\beta_c = \frac{1}{k \cdot \left( \frac{1-x}{2\mu_A} + \frac{x}{\alpha} + \frac{x}{\delta + \mu_I} \right)}. \quad (7.17)$$

Si nos vamos ahora al caso de una metapoblación con  $p = 0$ , bastará con calcular el valor de  $R_0$  mayor de todos los parches, aquel que tenga un número de contactos  $k_i$  mayor. Dicho valor será la suma de los contactos diurnos y nocturnos, ya que ocurren todos en el mismo nodo y viene dado por

$$k_i = z^D f_i + z^N \sigma_i. \quad (7.18)$$

En las siguientes secciones aclararemos y condensaremos todos estos conceptos en las ecuaciones de la dinámica para los formalismos de individuos tanto indistinguibles como distinguibles. Veremos también cómo es muy sencillo desarrollar por separado la teoría y las expresiones analíticas de un modelo compartimental (la parte que explica la enfermedad en sí misma, que es la que acabamos de estudiar) y luego unir las con los formalismos que explican los contactos con redes complejas (que es de cierta forma la *física* del sistema).

#### 7.4. Implementación del modelo en el formalismo de individuos indistinguibles

Vamos a describir el modelo para después derivar las ecuaciones markovianas que gobiernan la dinámica. Primero, los individuos decidirán si se desplazan o no de su nodo de origen con probabilidad  $p$ , y en caso de hacerlo, elegirán el parche al que viajar con la matriz origen-destino  $R_{ij}$ , siendo el peso de los enlaces la probabilidad de cada destino. Una vez han ocurrido los desplazamientos, ocurre el proceso de contagio. En él, los individuos tendrán una cantidad de contactos dada por  $k_i^D$  con otros agentes cualquiera que se encuentren en el mismo parche en ese momento. Esa cantidad vendrá dada por la función de contactos y la de normalización. Con cada uno de ellos tendrán una probabilidad de infectarse dependiendo del estado en el que se encuentren.

Después, vuelven a sus nodos de origen para llevar a cabo otro proceso de contagio, esta vez con los agentes que residan en el nodo. La cantidad de contactos vendrá dada por  $k_i^N$ , que depende del tamaño medio del hogar en cada parche. Al mismo tiempo, los individuos infectados tendrán una probabilidad de recuperarse (o pasar al siguiente compartimento). Un individuo que acaba de contagiarse no puede recuperarse en el mismo día.

Esta es la manera en la que se llevarán a cabo las simulaciones de agentes individuales, con un algoritmo de Monte Carlo. Queda ahora formalizar el modelo con las ecuaciones markovianas para el sistema mesoscópico. Las primeras ecuaciones consisten en aplicar las ya vistas en el modelo compartimental SEAPIDR (presentado en la sección 3.3.3), donde definimos una variable  $\rho^s(t)$  a cada nodo  $i$  (recordar que en el formalismo indistinguible, todos los individuos de un parche van a tener la misma probabilidad de encontrarse en cualquier estado).

$$\begin{aligned}
\rho_i^E(t+1) &= (1-\eta)\rho_i^E(t) + (1-\rho_i^E(t) - \rho_i^A(t) - \rho_i^P(t) - \rho_i^I(t) - \rho_i^D(t) - \rho_i^R(t))\Pi_i(t), \\
\rho_i^A(t+1) &= (1-\mu_A)\rho_i^A(t) + (1-x)\eta\rho_i^E(t), \\
\rho_i^P(t+1) &= (1-\alpha)\rho_i^P(t) + x\eta\rho_i^E(t), \\
\rho_i^I(t+1) &= (1-\delta-\mu_I)\rho_i^I(t) + \alpha\rho_i^P(t), \\
\rho_i^D(t+1) &= (1-\gamma)\rho_i^D(t) + \delta\rho_i^I(t), \\
\rho_i^R(t+1) &= \mu_A\rho_i^A(t) + \mu_I\rho_i^I(t) + \gamma\rho_i^D(t).
\end{aligned} \tag{7.19}$$

A continuación calculamos  $\Pi_i(t)$ , que tendrá en cuenta los lugares que un individuo puede visitar durante el día, y su residencia cuando ocurran los contactos nocturnos.

En el caso en que un agente decida no desplazarse (con probabilidad  $1-p$ ) tendrá una probabilidad  $P_i^D$  de contagiarse durante el día. En el caso con probabilidad  $1-P_i^D$  de que no se contagie, tendrá una probabilidad  $P_i^N$  de contagiarse por la noche.

Si decide viajar a  $j$  (con probabilidad  $pR_{ij}$ ) ocurre lo mismo pero con probabilidad  $P_j^D$  de contagiarse por el día. Sumando todos los casos posibles, tenemos

$$\Pi_i = (1-p) \left( P_i^D + (1-P_i^D)P_i^N \right) + p \left( \sum_{j=1}^N R_{ij} \left( P_j^D + (1-P_j^D)P_i^N \right) \right) \tag{7.20}$$

en cuyo caso  $P_i^D$  es la probabilidad de infectarse al encontrarnos en el nodo  $i$  por el día, que se obtiene a partir de la probabilidad de no contagiarse al contactar con un individuo, teniendo en cuenta la probabilidad de que esté infectado y en qué estado se encuentre.

Queda elevar todo eso al número de contactos que tendrá y calcular su complementaria.

$$P_i^D = 1 - \left( 1 - \frac{\sum_{s \in \{A, P, I\}} \beta_s (I_i^s)^{eff}}{n_i^{eff}} \right)^{z^D f_i}. \quad (7.21)$$

Lo mismo para  $P_i^N$ , excepto en este caso la proporción de infectados viene dada simplemente por  $\rho$ .

$$P_i^N = 1 - \left( 1 - \sum_{s \in \{A, P, I\}} \beta_s \rho_i^s \right)^{z^N \sigma_i}. \quad (7.22)$$

Calculamos la cantidad efectiva de infectados por el día teniendo en cuenta todos los agentes que viajan al nodo  $i$ , además de los que se quedan si residen en él. Lo mismo para la población efectiva.

$$(I_i^s)^{eff} = \sum_{j=1}^N \left( (1-p)\delta_{ij} + pR_{ji} \right) n_j \rho_j^s, \quad (7.23)$$

$$n_i^{eff} = \sum_{j=1}^N \left( (1-p)\delta_{ij} + pR_{ji} \right) \quad (7.24)$$

donde  $n_i$  es la población de cada nodo, la cual es conocida. Por último, recordar que  $f_i$  es la función de contactos que hemos definido como  $f_i = n_i^{eff} / a_i$ , siendo  $a_i$  el área del parche, que también conocemos, y  $\sigma$  el tamaño medio de los hogares. Las funciones de normalización son las ya definidas anteriormente como

$$z^D = \frac{\sum_{i=0}^N n_i^{eff} \langle k^D \rangle}{\sum_{i=0}^N n_i^{eff} f\left(\frac{n_i^{eff}}{a_i}\right)}, \quad z^N = \frac{\sum_{i=0}^N n_i \langle k^N \rangle}{\sum_{i=0}^N n_i \sigma_i}. \quad (7.25)$$

Notar que la función de normalización  $z^D$  puede cambiar dependiendo del valor de  $p$ . Esto se debe a que la cantidad total de contactos que se dan entre agentes puede cambiar dependiendo de su distribución. Si juntamos a todos los individuos en un parche, esperamos que la suma de los contactos que tienen todos los agentes crezca cuadráticamente, no linealmente como lo hace la función  $f_i$ . Como para  $p$  mayores esperamos una mayor concentración de personas en la mayoría de metapoblaciones, la cantidad de contactos total crecería si no fuese por la función de normalización  $z^D$ .

Esto significa que estamos forzando a la cantidad de contactos a ser constante con una  $z^D$  que será menor conforme aumentemos el valor de  $p$ , y veremos que contribuirá a un aumento del valor del umbral epidémico para valores de movilidad altos.

Es por esto que hay que ser cuidadosos con la forma en la que tomamos los datos y modelizamos la propagación de epidemias. Podemos obtener una cantidad concreta de los contactos que ocurren entre personas de observaciones en la vida real, pero si queremos predecir las consecuencias de cambiar la movilidad en una población, es lógico considerar si la media de contactos total se va a mantener igual o puede cambiar. Como queremos separar el estudio del efecto de la movilidad de otros factores como el distanciamiento social o aumento de la higiene como respuestas a una epidemia, tiene sentido forzar esa cantidad de contactos a ser la misma en cualquier caso.

#### 7.4.1. Verificación de las ecuaciones

Como cada vez que se obtienen ecuaciones para un nuevo modelo, debemos comprobar que predicen de forma correcta el resultado de las simulaciones Monte Carlo, las cuales recordamos que son mucho más costosas computacionalmente. Lo haremos de la misma forma que anteriormente, es decir, dibujando los diagramas de fase con los resultados de iterar las ecuaciones y mediando los valores obtenidos con varias simulaciones Monte Carlo.

Debido a que esta vez estamos utilizando un modelo cuyo compartimento final es el de recuperado (R), no podemos hacer la media sobre el tiempo en las simulaciones, ya que la cantidad de infectados finalmente siempre será cero. Por lo tanto, en el diagrama de fases representaremos la cantidad de individuos recuperados una vez la epidemia se ha extinguido, que es una manera de medir el impacto de la misma. Para obtener mejores resultados, haremos la media de muchas simulaciones. La evolución de un sistema de este tipo puede verse en la figura 7.1.

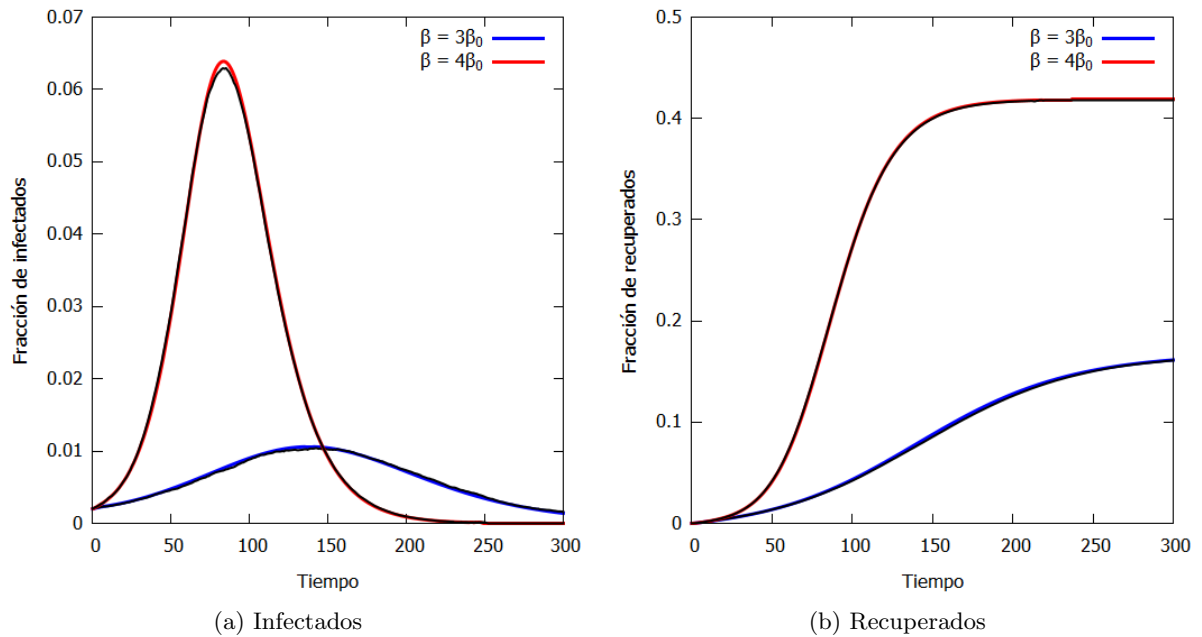


Figura 7.1: Simulación Monte Carlo (negro) frente a iteración de las ecuaciones markovianas (color) para el modelo SEAPIDR, en la figura de la izquierda se muestran los infectados en función del tiempo, y en la derecha la cantidad acumulada de recuperados.

Haciendo esto para distintos valores de  $p$  y  $\beta$ , tenemos los diagramas de fases de la figura 7.2, en los que vemos que efectivamente las ecuaciones describen correctamente el sistema. Si nos fijamos en los diagramas en concreto, vemos que la movilidad parece ser más detrimental al impacto que anteriormente. La diferencia es principalmente causada por la función de normalización, que como hemos comentado reducirá el número total de contactos para valores mayores de  $p$ , lo cual quiere decir que la movilidad por sí misma aparenta ser detrimental para valores altos de  $\beta$ .

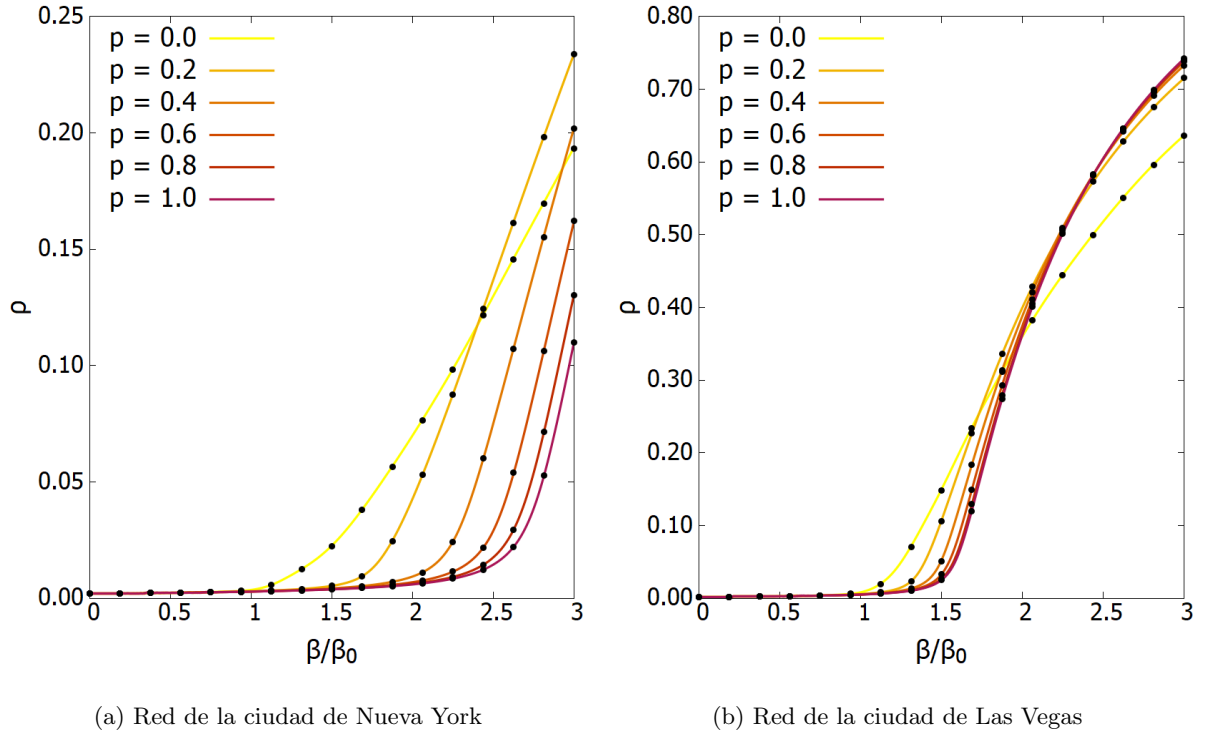


Figura 7.2: Diagrama de fases del modelo SEAPIDR para individuos indistinguibles. Las líneas continuas representan las iteraciones de las ecuaciones markovianas, y los puntos se corresponden a simulaciones Monte Carlo.

Vamos a obtener ahora las ecuaciones para el nuevo formalismo, para después ver cómo calcular los umbrales epidémicos de ambos.

### 7.5. Implementación del modelo en el formalismo de individuos distinguibles

Este formalismo es incluso más fácil de implementar en las simulaciones que el anterior. El funcionamiento general es exactamente el mismo, pero en este caso cada individuo tiene asignado un destino único posible, en vez de decidirlo aleatoriamente cada vez que vaya a viajar. Las ecuaciones markovianas para las transiciones de estados son las mismas que antes, con la diferencia en que cada variable tiene tantos componentes como enlaces  $L$  tiene la red, es decir que  $\rho_{ij}^s(t)$  nos da la fracción de individuos en el compartimento  $s \in \{E, A, P, I, D, R\}$  que residen en el nodo  $i$  y tienen como destino asignado el parche  $j$ .

$$\begin{aligned}
 \rho_{ij}^E(t+1) &= (1-\eta)\rho_{ij}^E(t) + (1-\rho_{ij}^E(t) - \rho_{ij}^A(t) - \rho_{ij}^P(t) - \rho_{ij}^I(t) - \rho_{ij}^D(t) - \rho_{ij}^R(t))\Pi_{ij}(t), \\
 \rho_{ij}^A(t+1) &= (1-\mu_A)\rho_{ij}^A(t) + (1-x)\eta\rho_{ij}^E(t), \\
 \rho_{ij}^P(t+1) &= (1-\alpha)\rho_{ij}^P(t) + x\eta\rho_{ij}^E(t), \\
 \rho_{ij}^I(t+1) &= (1-\delta-\mu_I)\rho_{ij}^I(t) + \alpha\rho_{ij}^P(t), \\
 \rho_{ij}^D(t+1) &= (1-\gamma)\rho_{ij}^D(t) + \delta\rho_{ij}^I(t), \\
 \rho_{ij}^R(t+1) &= \mu_A\rho_{ij}^A(t) + \mu_I\rho_{ij}^I(t) + \gamma\rho_{ij}^D(t).
 \end{aligned} \tag{7.26}$$

Para calcular  $\Pi_{ij}$  usamos la misma lógica. Los individuos que no se desplacen, con probabilidad  $1-p$ , tienen una probabilidad  $P_i^D$  de infectarse por el día, y en caso de no hacerlo (caso  $1-P_i^D$ ), pueden contraer la enfermedad por la noche con probabilidad  $P_i^N$ . Los que se desplacen con probabilidad  $p$ , tienen una probabilidad  $P_j^D$  de contagiarse por el día en el nodo  $j$ , y de nuevo

$P_i^N$  de contagiarse por la noche en su residencia en caso de no haberlo hecho durante el día, resultando en

$$\Pi_{ij} = (1 - p) \left( P_i^D + (1 - P_i^D) P_i^N \right) + p \left( P_j^D + (1 - P_j^D) P_i^N \right) \quad (7.27)$$

donde las probabilidades de contagiarse en cada nodo son iguales a las del formalismo de individuos indistinguibles, excepto que hay que contar la cantidad de infectados en el nodo  $i$  para calcular  $P_i^N$  como suma de los diferentes  $\rho_{ij}^s$ .

$$P_i^D = 1 - \left( 1 - \frac{\sum_{s \in \{A, P, I\}} \beta_s (I_i^s)^{eff}}{n_i^{eff}} \right)^{z^D f_i}, \quad (7.28)$$

$$P_i^N = 1 - \left( 1 - \sum_{s \in \{A, P, I\}} \beta_s \frac{\sum_{j=1}^N n_{ij} \rho_{ij}^s}{\sum_{j=1}^N n_{ij}} \right)^{z^N \sigma_i}. \quad (7.29)$$

Queda contar la población efectiva y la cantidad de infectados en cada nodo, a partir de la matriz origen-destino de este formalismo, es decir,  $n_{ij}$ .

$$n_i^{eff} = \sum_{j=1}^N p n_{ji} + \sum_{j=1}^N (1 - p) n_{ij}, \quad (7.30)$$

$$(I_i^s)^{eff} = \sum_{j=1}^N p n_{ji} \rho_{ji}^s + \sum_{j=1}^N (1 - p) n_{ij} \rho_{ij}^s, \quad s \in \{A, P, I\}. \quad (7.31)$$

Por último, el valor de la función de contactos será  $f_i = n_i^{eff} / a_i$ ,  $\sigma_i$  el tamaño del hogar, que no afecta al resultado si no hay datos concretos para cada parche, y las funciones de normalización son

$$z^D = \frac{\sum_{i=0}^N n_i^{eff} \langle k^D \rangle}{\sum_{i=0}^N n_i^{eff} f\left(\frac{n_i^{eff}}{a_i}\right)}, \quad z^N = \frac{\sum_{i=0}^N n_i \langle k^N \rangle}{\sum_{i=0}^N n_i \sigma_i}. \quad (7.32)$$

### 7.5.1. Verificación de las ecuaciones

Dibujaremos los diagramas de fase para este modelo con el mismo procedimiento que hemos utilizado en la sección anterior, es decir, haciendo la media de muchas simulaciones Monte Carlo y compararla con la iteración de las ecuaciones markovianas. El resultado se muestra en la figura 7.3, donde otra vez vemos que las ecuaciones son correctas.

Si nos fijamos en las curvas representadas, son bastante distintas respecto a las del formalismo de individuos indistinguibles, incluso habiendo realizado las simulaciones para las mismas redes (Nueva York y Las Vegas). Aquí encontramos que el umbral epidémico para  $p$  mayor disminuye drásticamente, a pesar del detrimento inicial, al contrario que en las figuras 7.2, donde el umbral sigue aumentando con  $p$ . Nos centraremos en este aspecto cuando obtengamos las expresiones para el umbral epidémico. Observamos también que llega un punto en el que el impacto para los distintos valores de  $p$  se invierte, y la  $p$  con mayor umbral epidémico pasa a ser la que más infecciones va a permitir.

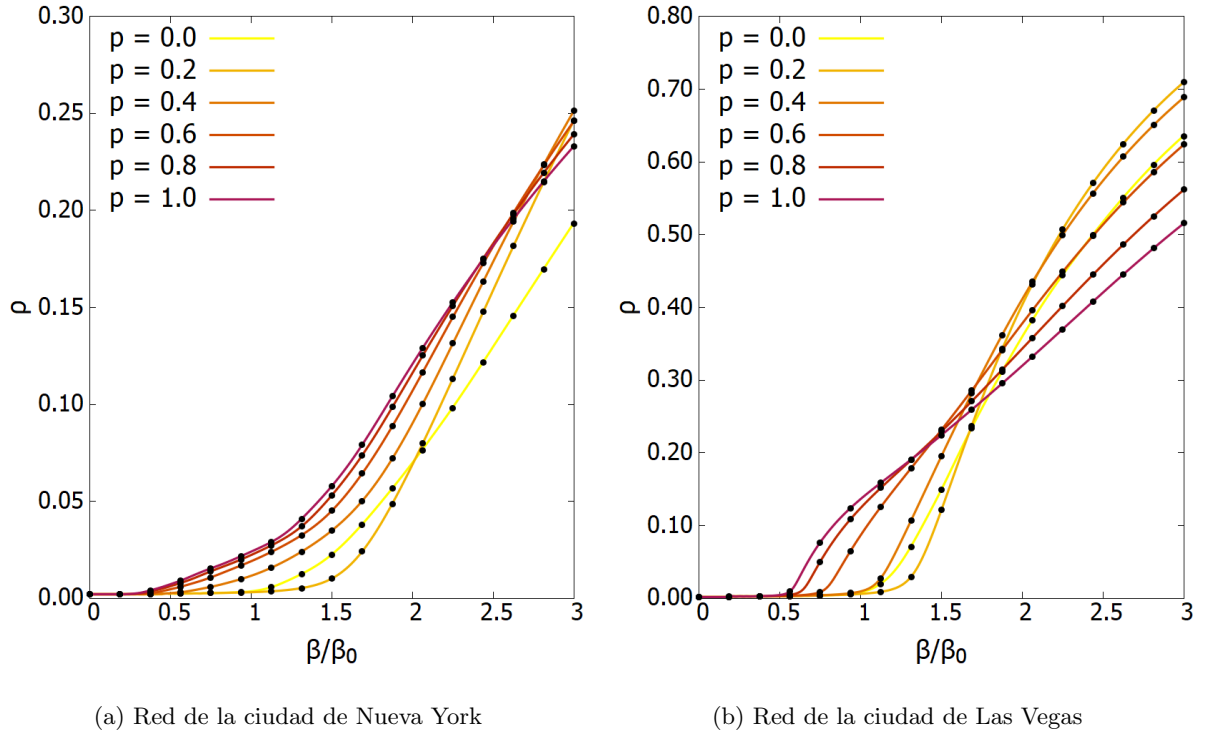


Figura 7.3: Diagrama de fases del modelo SEAPIDR para individuos distinguibles. Las líneas continuas representan las iteraciones de las ecuaciones markovianas, y los puntos se corresponden a simulaciones Monte Carlo.

De nuevo, todo se reduce a la heterogeneidad del sistema, que sabemos que es mayor en el formalismo distinguible. A pesar de que mantener concentrados a los individuos en una localización reduce el umbral epidémico, la situación es diferente cuando la probabilidad de contagio  $\beta$  es tan alta que el problema no es que no se pueda mantener la epidemia activa, sino que se propague suficientemente rápido antes de que se inmunice la población.

## 7.6. Cálculo del umbral epidémico con metapoblaciones

En esta sección vamos a ver cómo gracias al trabajo realizado en la sección 7.3.2 tenemos la mayor parte del trabajo hecho. Para empezar, las transiciones entre estados son las mismas, excepto que cada nodo (en el formalismo indistinguible) o enlace (en el distinguible) tiene asociado su propia componente. Por lo tanto, la cantidad de ecuaciones será igual a la anterior multiplicada por  $N$  o por  $N^2$ . La matriz  $\Sigma$  será multiplicada por la matriz identidad (producto de Kronecker) porque los estados no transicionan de un parche o enlace a otro.

$$\Sigma = \mathbf{I} \otimes \begin{pmatrix} -\eta & 0 & 0 & 0 \\ (1-x)\eta & -\mu_A & 0 & 0 \\ x\eta & 0 & -\alpha & 0 \\ 0 & 0 & \alpha & -(\delta + \mu_I) \end{pmatrix}. \quad (7.33)$$

La matriz identidad será de tamaño  $N \times N$  en el formalismo indistinguible y  $N^2 \times N^2$  en el distinguible.

Para la matriz de transmisión  $\mathbf{T}$  tenemos primero que linealizar las ecuaciones de ambos formalismos.

### 7.6.1. Linealización de las ecuaciones del formalismo indistinguible

Empezamos por linealizar las ecuaciones de las probabilidades de contagio (7.21), asumiendo  $\rho_i^s(t) \approx \varepsilon_i^s \ll 1$ , para todos los nodos.

$$\begin{aligned}
P_i^D &= 1 - \left( 1 - \frac{\sum_{s \in \{A, P, I\}} \beta_s(I_i^s)^{eff}}{n_i^{eff}} \right)^{z^D f_i} \approx z^D f_i \frac{\sum_{s \in \{A, P, I\}} \beta_s(I_i^s)^{eff}}{n_i^{eff}} = \\
&= \frac{z^D f_i}{n_i^{eff}} \sum_{j=1}^N \left( (1-p)\delta_{ij} + pR_{ji} \right) n_j \sum_{s \in \{A, P, I\}} \beta_s \varepsilon_j^s, \\
P_i^N &= z^N \sigma_i \sum_{s \in \{A, P, I\}} \beta_s \varepsilon_i^s.
\end{aligned}$$

Con esto podemos calcular el valor linealizado de  $\Pi_i$ , que primero tenemos que escribir como

$$\begin{aligned}
\Pi_i &= (1-p) \left( P_i^D + (1-P_i^D)P_i^N \right) + p \left( \sum_{j=1}^N R_{ij} \left( P_j^D + (1-P_j^D)P_j^N \right) \right) \approx \\
&= (1-p) \left( P_i^D + P_i^N \right) + p \left( \sum_{j=1}^N R_{ij} \left( P_j^D + P_j^N \right) \right) = (1-p)P_i^D + P_i^N + p \sum_{j=1}^N R_{ij}P_j^D. \quad (7.34)
\end{aligned}$$

Sustituyendo y reordenando los términos llegamos a la expresión de  $\Pi_i$ .

$$\begin{aligned}
\Pi_i &= \sum_{j=1}^N \left[ \left( (1-p)^2 \frac{z^D f_i}{n_i^{eff}} + \frac{z^N \sigma_i}{n_i} \right) \delta_{ij} + p(1-p) \left( R_{ji} \frac{z^D f_i}{n_i^{eff}} + R_{ij} \frac{z^D f_j}{n_j^{eff}} \right) + \right. \\
&\quad \left. p^2 \sum_{l=1}^N R_{il} R_{jl} \frac{z^D f_l}{n_l^{eff}} \right] n_j \sum_{s \in \{A, P, I\}} \beta_s \varepsilon_j^s. \quad (7.35)
\end{aligned}$$

Esta ecuación obviamente se puede escribir de forma matricial, donde la matriz en concreto tendrá componentes  $i, j$  y  $s$  porque tiene en cuenta las interacciones entre diferentes compartimentos. En realidad, todos los valores de  $\Pi_i$  hacen referencia a la transmisión al compartimento de expuestos (E), de la misma forma que en la ecuación (7.11). Esta vez, esa matriz tiene en cuenta los contactos entre distintos parches de la metapoblación, pero se puede construir fácilmente separando los compartimentos (el sumatorio en  $s$ ) y los parches (el sumatorio en  $j$ ), que vemos que en la expresión anterior son independientes. La matriz de mezcla es

$$\begin{aligned}
M_{ij} &= \left[ \left( (1-p)^2 \frac{z^D f_i}{n_i^{eff}} + \frac{z^N \sigma_i}{n_i} \right) \delta_{ij} + p(1-p) \left( R_{ji} \frac{z^D f_i}{n_i^{eff}} + R_{ij} \frac{z^D f_j}{n_j^{eff}} \right) + \right. \\
&\quad \left. p^2 \sum_{l=1}^N R_{il} R_{jl} \frac{z^D f_l}{n_l^{eff}} \right] n_j \quad (7.36)
\end{aligned}$$

y con esto la matriz de transmisión  $\mathbf{T}$  se puede escribir como el producto de Kronecker de la



matriz de mezcla y la matriz de transmisión original.

$$\mathbf{T} = \mathbf{M} \otimes \begin{pmatrix} 0 & \beta_A & \beta_P & \beta_I \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (7.37)$$

Vemos la similitud con la ecuación (7.11), donde el valor de  $k$  es ahora una matriz que relaciona los distintos parches de la metapoblación. Vamos a poder hacer esta separación siempre que los compartimentos se comporten de forma independiente a los nodos o enlaces de la red. Antes de seguir, vamos a obtener la matriz de mezcla para el formalismo distinguible.

### 7.6.2. Linealización de las ecuaciones del formalismo distinguible

Teniendo en cuenta que las variables se aplican a cada enlace de la red, y asumiendo  $\rho_{ij}^s \approx \varepsilon_{ij}^s \ll 1$ , podemos linealizar las ecuación 7.28 para conocer las probabilidades de infección

$$P_i^D = z^D f_i \frac{\sum_{s \in \{A,P,I\}} \beta_s(I_i^{eff})}{n_i^{eff}} \approx p \frac{z^D f_i}{n_i^{eff}} \sum_{j=1}^N n_{ji} \sum_{s \in \{A,P,I\}} \beta_s \varepsilon_{ji}^s + (1-p) \frac{z^D f_i}{n_i^{eff}} \sum_{j=1}^N n_{ij} \sum_{s \in \{A,P,I\}} \beta_s \varepsilon_{ij}^s, \quad (7.38)$$

$$P_i^N = \frac{z^N \sigma_i}{n_i} \sum_{j=1}^N n_{ij} \sum_{s \in \{A,P,I\}} \varepsilon_{ij}^s. \quad (7.39)$$

Operando y cambiando ciertos índices de los sumatorios para que sean coherentes, tenemos la expresión para  $\Pi_{ij}$  linealizado

$$\begin{aligned} \Pi_{ij} = \sum_{k,l=1}^N \left[ (1-p) p \frac{z^D f_k}{n_k^{eff}} n_{lk} \delta_{ik} + (1-p)^2 \frac{z^D f_l}{n_l^{eff}} n_{lk} \delta_{il} + p^2 \frac{z^D f_k}{n_k^{eff}} n_{lk} \delta_{jk} + \right. \\ \left. + p(1-p) \frac{z^D f_l}{n_l^{eff}} n_{lk} \delta_{jl} + \frac{z^N \sigma_l}{n_l} n_{lk} \delta_{il} \right] \beta_s \varepsilon_{lk}^s \quad (7.40) \end{aligned}$$

donde, si lo escribimos de forma matricial, la matriz de mezcla es

$$\begin{aligned} M_{jk}^{il} = \left( (1-p)^2 \frac{z^D f_l}{n_l^{eff}} + \frac{z^N \sigma_l}{n_l} n_{lk} \delta_{il} \right) n_{lk} \delta_{il} + (1-p) p \frac{z^D f_k}{n_k^{eff}} n_{lk} \delta_{ik} + \\ + p(1-p) \frac{z^D f_l}{n_l^{eff}} n_{lk} \delta_{jl} + p^2 \frac{z^D f_k}{n_k^{eff}} n_{lk} \delta_{jk}. \quad (7.41) \end{aligned}$$

De esta manera, en este formalismo también se puede escribir la matriz de transmisión como en (7.37). Usaremos las propiedades del producto de Kronecker [25] para obtener el valor de  $R_0$  de forma sencilla. Primero, dado que el producto de Kronecker cumple que si  $A, B$  son matrices no singulares, entonces  $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ , por lo que podemos decir de la matriz de transición en (7.33) lo siguiente:

$$\Sigma^{-1} = \mathbf{I}^{-1} \otimes \begin{pmatrix} -\eta & 0 & 0 & 0 \\ (1-x)\eta & -\mu_A & 0 & 0 \\ x\eta & 0 & -\alpha & 0 \\ 0 & 0 & \alpha & -(\delta + \mu_I) \end{pmatrix}^{-1} = \mathbf{I} \otimes \begin{pmatrix} \frac{1}{\eta} & 0 & 0 & 0 \\ \frac{1-x}{\mu_A} & \frac{1}{\mu_A} & 0 & 0 \\ \frac{x}{\alpha} & 0 & \frac{1}{\alpha} & 0 \\ \frac{x}{\delta + \mu_I} & 0 & \frac{1}{\delta + \mu_I} & \frac{1}{\delta + \mu_I} \end{pmatrix}. \quad (7.42)$$

Queda por último el producto  $-\mathbf{T}\Sigma^{-1}$ , para el que usamos la propiedad siguiente: sea  $A \in M_{m,n}$ ,  $B \in M_{p,q}$ ,  $C \in M_{n,k}$  y  $D \in M_{q,r}$ , entonces  $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$ . Si aplicamos esto a las matrices  $\mathbf{T}$  y  $\Sigma^{-1}$ , que son producto tensorial de las originales con la identidad y la matriz de mezcla  $\mathbf{M}$  respectivamente, tenemos a la izquierda simplemente la matriz de mezcla (solamente se multiplica por la identidad) y a la derecha el producto original visto en (7.14). El resultado es:

$$\mathbf{K}_L = -\mathbf{T}\Sigma^{-1} = \mathbf{M} \otimes \begin{pmatrix} \beta_A \frac{1-x}{\mu_A} + \beta_P \frac{x}{\alpha} + \beta_I \frac{x}{\delta + \mu_I} & \frac{\beta_P}{\mu_A} & \frac{\beta_P}{\alpha} + \frac{\beta_I}{\delta + \mu_I} & \frac{\beta_I}{\delta + \mu_I} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (7.43)$$

Queremos obtener los autovalores de este producto de Kronecker, que por otra propiedad serán los productos de los autovalores de las dos originales, y están asociados a los productos de Kronecker de sus autovectores. Por lo tanto, como sólo hay un autovalor distinto de cero en la matriz asociada a los compartimentos, el mayor será

$$R_0 = \Lambda_{\max}(\mathbf{M}) \cdot \left( \beta_A \frac{1-x}{\mu_A} + \beta_P \frac{x}{\alpha} + \beta_I \frac{x}{\delta + \mu_I} \right). \quad (7.44)$$

Si ahora suponemos que  $\beta \equiv 2\beta_A = \beta_P = \beta_I$ , y tomando el caso en que  $R_0 = 1$ , obtenemos el umbral epidémico

$$\beta_c = \frac{1}{\Lambda_{\max}(\mathbf{M}) \cdot \left( \frac{1-x}{2\mu_A} + \frac{x}{\alpha} + \frac{x}{\delta + \mu_I} \right)}. \quad (7.45)$$

Por lo tanto, a pesar de estar trabajando con un modelo considerablemente más complejo que los anteriores, el cálculo del umbral epidémico se reduce de nuevo a obtener el máximo autovalor de la matriz de mezcla. Recordamos que esta matriz tiene tamaño  $N \times N$  en el formalismo de individuos indistinguibles, y  $N^2 \times N^2$  en el de individuos distinguibles, pero se puede optimizar de la misma forma que como hemos visto en la sección 5.2, porque las componentes que son distintas de cero son las mismas, a pesar de tener valores diferentes.

Vamos a crear los mapas de calor para ver el impacto de la epidemia en función de la movilidad  $p$  y la probabilidad de contagio  $\beta$ , y comprobar que la expresión del umbral epidémico es correcta.

Los mapas de calor en el modelo SEAPIDR se crean iterando las ecuaciones markovianas a partir de una semilla de infectados, hasta que la epidemia se extingue y contamos los recuperados al final, que representan el impacto. Recordar que las ecuaciones markovianas son deterministas, por lo que únicamente es necesario realizar una iteración cada vez. Representamos los resultados en la figura 7.4.

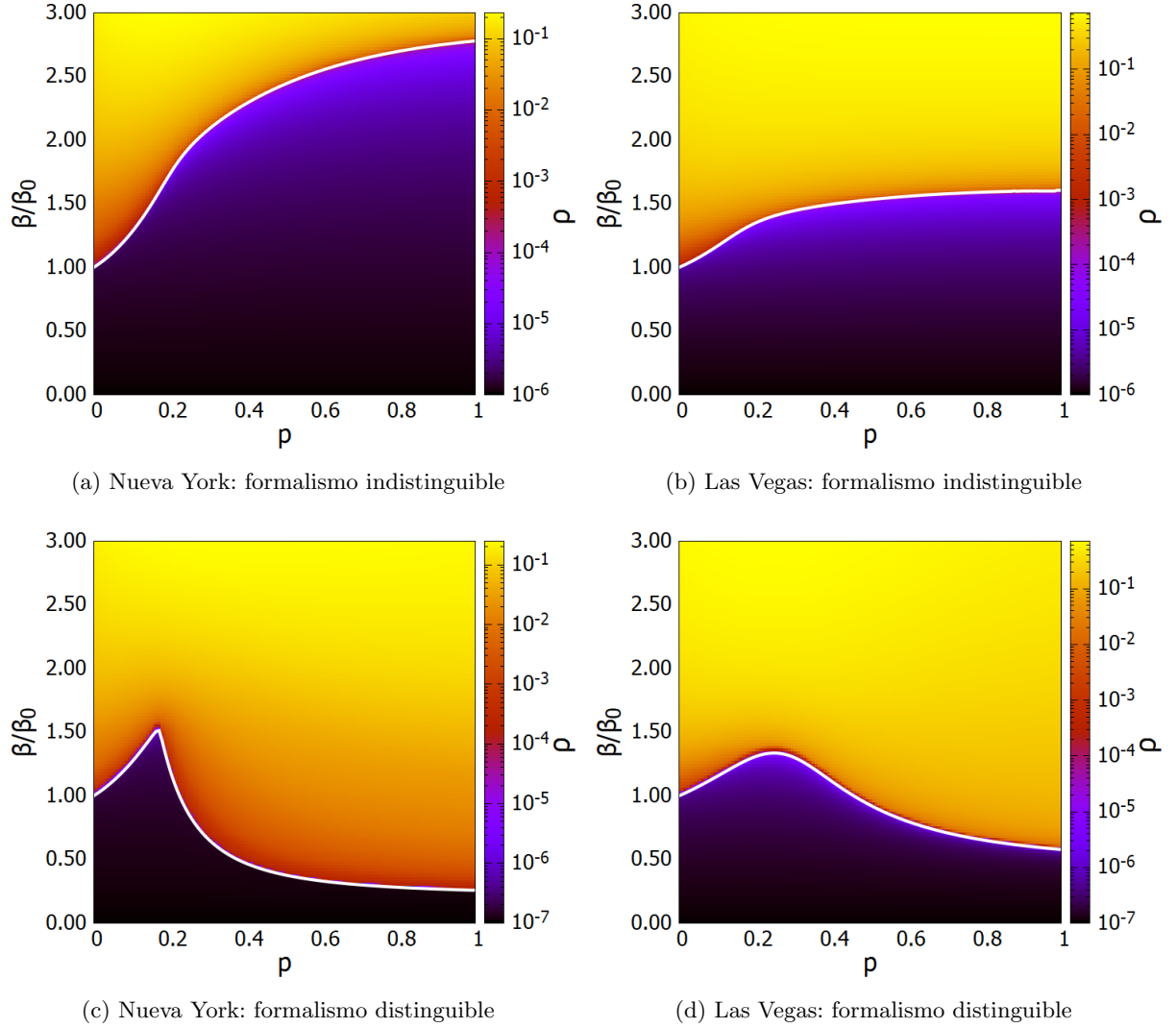


Figura 7.4: Mapas de calor del impacto de la epidemia (individuos recuperados) en función de la movilidad (eje horizontal) y la probabilidad de contagio (eje vertical), a partir de iterar las ecuaciones markovianas. La línea blanca representa el umbral epidémico  $\beta_c(p)$ . En las figuras de arriba se ha usado el formalismo de individuos indistinguibles, y en las de abajo el de individuos distinguibles. Las redes se corresponden a las ciudades de Nueva York y Las Vegas.

Algo que aparece por primera vez es un detrimento continuado en el formalismo de individuos indistinguibles, donde el umbral epidémico crece hasta llegar a  $p = 1$ . Mientras esto es algo que esperábamos al definir la función de normalización de forma que generalmente disminuyese con la movilidad, es notable el hecho de que esto no ocurra en absoluto para el formalismo de individuos distinguibles. De hecho, esta figura demuestra la gran importancia del hecho de considerar los viajes recurrentes, y diferenciar a los agentes según sus patrones específicos de movilidad.

La comparación entre los umbrales epidémicos de ambos formalismos revela cómo el formalismo distinguible puede ser capaz de explicar brotes epidémicos en escenarios no contemplados por el indistinguible, incluso para valores de la probabilidad de contagio  $\beta$  mucho mayores (hasta diez veces mayor en la red de Nueva York). Ya hemos visto en la sección 6 cómo usar la expresión del umbral epidémico para llevar a cabo medidas de restricción de movilidad que reduzcan la propagación de la epidemia considerablemente. A continuación vamos a ver cómo podemos utilizar el mismo principio para optimizar los recursos de diagnóstico y predecir los lugares más susceptibles de brotes epidémicos.

## 8. Flujos críticos para el diagnóstico de brotes epidémicos

El máximo autovalor de la matriz de mezcla está asociado a un vector propio de la misma, que se corresponde con las proporciones de infectados que vamos a encontrar en los enlaces de la red cuando la epidemia tenga un  $R_0$  cercano a 1. Además, esta información nos puede ser muy útil ya que podemos extrapolarla a cualquier enfermedad: recordar que la matriz está construida exclusivamente a partir de datos de la metapoblación.

Vamos entonces a poner a prueba la capacidad de predicción del vector propio obtenido por el método que hemos comentado para probabilidades de contagio que excedan considerablemente el valor del umbral epidémico. El objetivo es demostrar que concentrando los recursos de diagnóstico de una enfermedad en los enlaces críticos para la supervivencia de una epidemia, podemos detectar los brotes antes y con positividad mucho mayor que si distribuimos las pruebas de forma aleatoria en toda la población.

El algoritmo consistirá en realizar simulaciones Monte Carlo siguiendo el modelo descrito en la sección 7, y realizar un número determinado de pruebas de distintas formas. La primera será hacerlas aleatoriamente y de forma uniforme en toda la red. La segunda se centrará en los enlaces cuyas componentes en el vector propio sean las mayores. Podemos variar la cantidad de enlaces que incluimos en el diagnóstico.

Representamos la media de la cantidad de infectados detectados en cada paso temporal entre varias simulaciones Monte Carlo, donde se han realizado un total de 10 mil pruebas. La diferencia se encuentra en cuántos enlaces escogemos para centrar los tests en sus agentes. Representamos los resultados en la figura 8.1.

Estos resultados nos muestran que es posible aumentar la capacidad de detectar de forma temprana un brote epidémico en una población, y obtener una tasa de positividad mucho mayor, si centramos los recursos en los flujos de movilidad más críticos de la red siguiendo el procedimiento que hemos descrito. Esto puede ser muy beneficioso a la hora de responder más rápidamente a una epidemia y de que las autoridades se vean inclinadas a tomar medidas con antelación. Por ejemplo, durante el brote de COVID-19 en España, se estableció un procedimiento por el cual las restricciones se harían más estrictas dependiendo de la incidencia detectada en el momento, por lo que una mayor positividad detectada puede suponer un impacto considerablemente menor de la enfermedad.

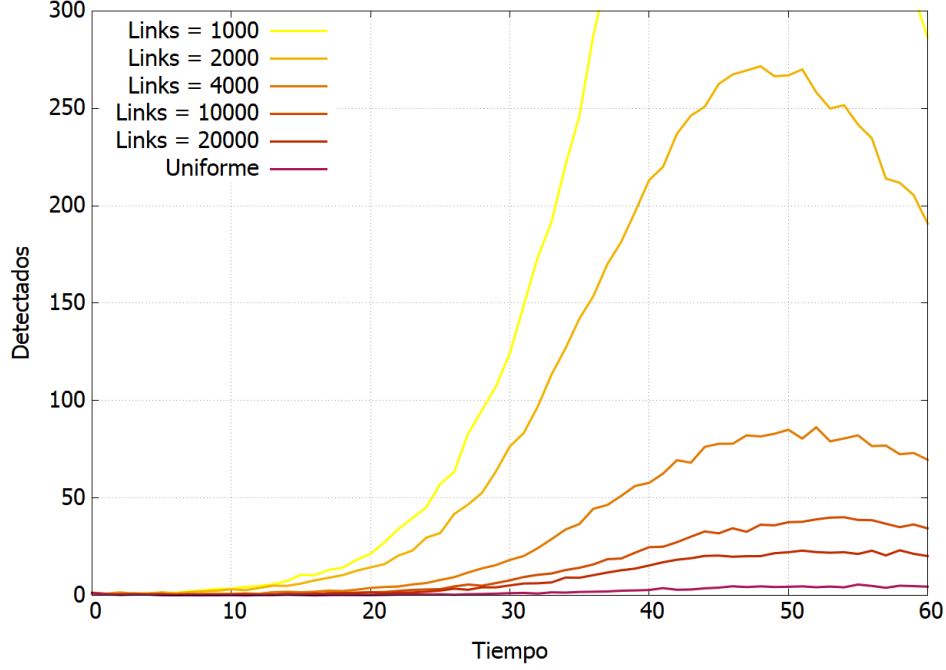


Figura 8.1: Cantidad de detectados de una muestra de 10 mil agentes aleatorios, escogidos de entre los  $L$  enlaces con mayor componente del vector propio asociado al mayor autovalor de la matriz de mezcla. La muestra *uniforme* es recogida por igual en toda la red, que consta de más de 147 mil enlaces. La metapoblación se corresponde a la ciudad de Nueva York. El valor de  $\beta$  es tres veces el valor de  $\beta_c$  para una movilidad  $p = 0,7$ , por lo que  $R_0 = 3$ .

## 9. Aplicación al caso del estado de Nueva York: diferentes escalas espaciales

Para cerrar el trabajo, vamos a abordar un problema con el que no hemos lidiado hasta ahora y del cual no conocemos sus consecuencias. Se trata de la escala espacial con la que seleccionamos nuestras subpoblaciones y el efecto que puede tener en los resultados. La razón por la que estamos estudiando la movilidad a través de la teoría de metapoblaciones es para evitar la hipótesis de campo medio. Sin embargo, asumir una escala demasiado grande para la resolución de los parches puede suponer la asunción de campo medio en lugares donde sea incorrecto hacerlo. Para comprobarlo, refinaremos parámetros del último modelo desarrollado (sección 7.5) para ponernos en el caso más realista posible y realizaremos las simulaciones para el estado de Nueva York, teniendo en cuenta las distintas resoluciones espaciales para la definición de parche y esto nos permitirá comparar los resultados en las distintas resoluciones.

Durante la última parte del trabajo hemos utilizado una función lineal para describir la cantidad de contactos que se dan en una población en función de su densidad. A pesar de que es una hipótesis intuitiva en principio<sup>2</sup> y nos permite resaltar las heterogeneidades en los sistemas, si queremos aplicarlo a un caso real y ser lo más exactos que podamos debemos elegir una más adecuada.

<sup>2</sup>Si consideramos un radio constante alrededor de cada individuo, donde tendrá un contacto con cada persona que entre en su área, la cantidad de contactos crecerá de forma lineal con la densidad de población.

La función que vamos a utilizar fue escogida en el modelo desarrollado en [21] en el que se estudia la propagación del COVID-19 en España, y viene descrita en [26]. Viene dada por

$$f(\rho) = 2 - e^{-\alpha\rho} \quad (9.1)$$

donde  $\rho$  es la densidad y  $\alpha$  es un factor basado en la media de densidades de la metapoblación para tener una heterogeneidad mayor. De esta forma el factor por el que multiplicamos la cantidad de contactos estará acotado tal que  $f \in [1, 2)$ , poniendo límites inferior y superior. Esta función va a suponer una homogeneización muy importante de toda la red, donde el foco de una epidemia no va a estar completamente localizado en el parche más denso sino que se observarán brotes más dispersos, tal y como ocurre en los casos reales. Al mismo tiempo esto causará que no veamos claras diferencias en las variables macroscópicas al variar la movilidad o cambiar de escala.

Otra sutileza que tendremos en cuenta tiene que ver con el valor de la función de normalización  $z^D$  para todas las escalas, escogiendo aquella que fija la media de contactos total para la escala más pequeña. Hacemos esto para que todas las simulaciones sigan las mismas condiciones impuestas cuando se calculen los contactos en cada iteración.

Escogemos tres escalas diferentes que representan subpoblaciones en el estado de Nueva York. La primera es la de códigos postales, que es la que hemos estado utilizando a lo largo del trabajo y es la más detallada. Existen 1768 códigos postales diferentes en Nueva York. La segunda será la de subcondados, que es ligeramente mayor, con 1012 de ellos. Finalmente, tenemos la mayor de todas, los condados, de los que únicamente hay 62. En algunas zonas los subcondados se repartirán de forma similar a los códigos postales, mientras que en otros sitios como en la ciudad de Nueva York los condados y subcondados tienen una relación de uno a uno. En los resultados a continuación omitiremos la resolución a escala de subcondados por quedar sobrepuesta con la de códigos postales.

Lo primero que vamos a analizar son los diagramas de fase y los umbrales epidémicos, que representamos en las figuras en 9.1. Estos resultados presentan diferencias muy pequeñas con la movilidad en comparación con anteriormente, cuando utilizábamos una función de contactos lineal. Sin embargo, aún podemos observar el comportamiento esperado. Ya que la red de códigos postales es más precisa, será más heterogénea y su umbral epidémico es menor. Sin embargo, en el régimen supercrítico las redes homogéneas tienden a facilitar más la propagación de una epidemia y las curvas del diagrama de fases acaban cruzándose.

El resultado más interesante lo obtenemos cuando comparamos el impacto en las diferentes regiones de la red. Si realizamos una simulación para cada escala espacial, con una probabilidad de contagio tal que  $R_0 = 1.4$  (del orden de la gripe estacional), podemos observar cómo las predicciones cambian considerablemente dependiendo de la escala utilizada. Mostramos los resultados en la figura 9.2.

Estos resultados son de gran importancia, porque demuestran varias cosas. Primero, a pesar de que los resultados macroscópicos (impacto a escala global) coincidan, esto no es un indicador completo sobre lo que está ocurriendo en el sistema. Segundo, la escala a la que se realizan las simulaciones es crucial y aumentar su resolución va a suponer una mejora considerable en su capacidad de simular procesos de contagio realistas. Por supuesto, una epidemia no va a estar focalizada por completo en la ciudad de Nueva York para después no propagarse por zonas menos densas, sino que siempre van a existir pequeñas zonas muy pobladas en condados que de media tienen menos densidad, donde los contagios van a tener un efecto importante. Agregar diferentes códigos postales en un único condado lleva a distribuciones de densidad más

homogéneas, lo que acaba evitando brotes de la enfermedad en localizaciones que podrían haber sido focos a una resolución más detallada, por lo que el impacto en la escala de condados en muchas regiones es mucho menor.

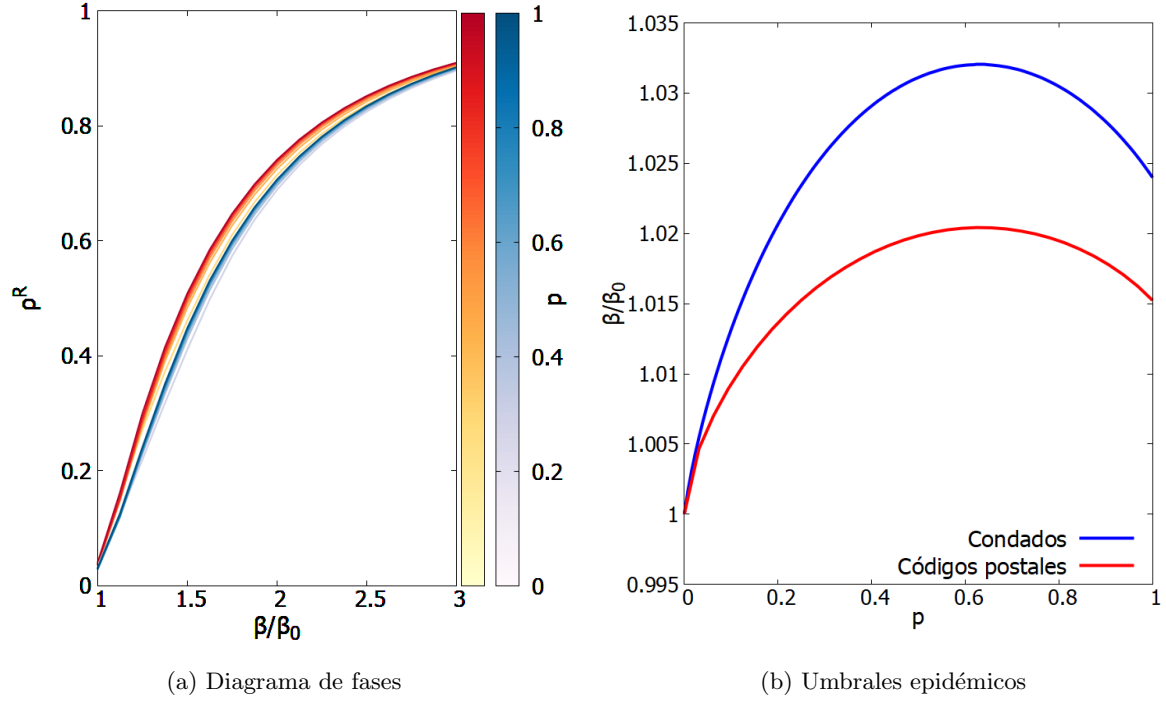


Figura 9.1: Comparaciones de los resultados de simulaciones para las escalas de códigos postales (rojos) y condados del estado de Nueva York (azules). A la izquierda, el diagrama de fases en el que se representan la proporción de individuos recuperados de la epidemia, que representa el impacto en función de la infectividad del patógeno  $\beta$ , y a la derecha las curvas de los umbrales epidémicos en función de la probabilidad de desplazamiento.

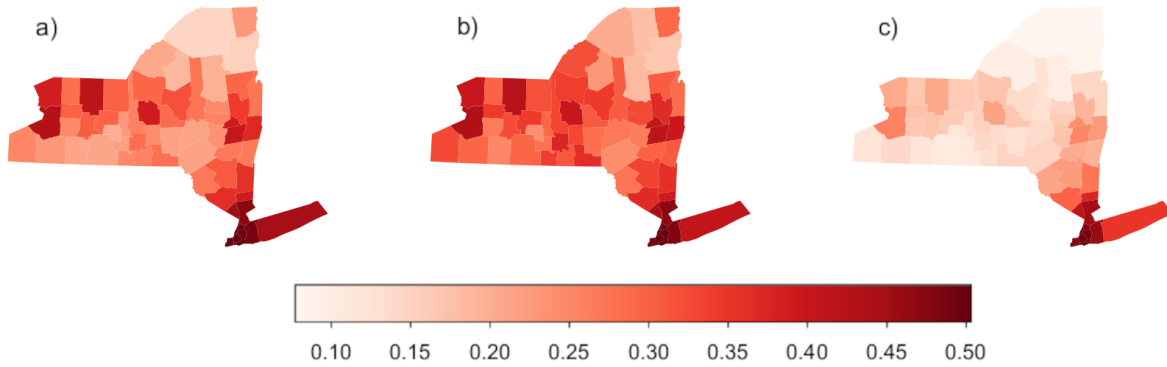


Figura 9.2: Proporción de recuperados (impacto de la epidemia) para a) escala a nivel de código postal, b) escala a nivel de subcondado, c) escala a nivel de condado. Los mapas se representan a nivel de condado, donde la cantidad de recuperados se ha sumado entre todos los códigos postales o subcondados que pertenecen a cada uno.

## 10. Conclusiones

A lo largo del trabajo hemos presentado el marco existente en el campo de modelización de la propagación de epidemias, el cual ha visto un gran crecimiento durante los últimos años, con una considerable cantidad de artículos siendo publicados al respecto año tras año. Darnos cuenta de cómo unir el campo de la Física de Sistemas Complejos con la epidemiología puede suponer un salto enorme en la capacidad predictiva de estos modelos nos ha llevado a continuar los trabajos realizados previamente, con el objetivo de mejorar los formalismos y demostrar las diversas formas en las que podemos aplicarlos para la prevención y mitigación de estos fenómenos. No solamente hemos introducido un formalismo distinto a la hora de tratar la caracterización de los distintos grupos de agentes en el sistema, sino también otros detalles que suponen una mejora considerable del realismo del modelo (centrándonos en cómo las concentraciones de personas influyen en la cantidad de contactos, y cómo estos se llevan a cabo). A cada paso hemos utilizado las simulaciones de agentes individuales que se querían describir para comprobar y verificar las ecuaciones desarrolladas, al mismo tiempo que tratamos de interpretar todos estos resultados.

Se ha mostrado cómo la movilidad juega un papel crucial en la propagación de enfermedades en una población, ya sea facilitando su contagio en el régimen supercrítico como suponiendo un detrimento en el umbral para algunos casos. Hemos comprobado cómo cambiar la naturaleza de los desplazamientos descritos en el formalismo, incluso de forma sutil, puede alterar considerablemente el impacto de una epidemia. En concreto se ve cómo el formalismo de individuos distinguibles, en el que los viajes ocurren de forma recurrente, resulta en umbrales epidémicos considerablemente menores que el anterior.

Describir la dinámica de estos sistemas con ecuaciones de Markov en tiempo discreto nos ha permitido también obtener resultados analíticos, tales como la matriz de mezcla (que representa los contactos que ocurren entre agentes de cualquier lugar de la red), la expresión del número básico de reproducción  $R_0$  y el umbral epidémico, además del estado fundamental en el que se encuentra la infección cerca del umbral epidémico cuando calculamos el vector propio asociado al mayor autovalor. Hemos encontrado una manera de separar la parte del modelo que explica los compartimentos característicos de la enfermedad frente al formalismo utilizado para describir los contactos y la topología de la red, para poder aplicar el desarrollo a diferentes modelos de forma inmediata.

Estos resultados analíticos nos han permitido poner a prueba políticas de control sobre la epidemia, ya sea restringiendo la movilidad entre subpoblaciones o centrando el diagnóstico de contagios en los flujos críticos de la red, para poder responder de forma más efectiva con mayor rapidez.

Por último, los resultados obtenidos al comparar las diferentes escalas espaciales a la hora de realizar las simulaciones nos han mostrado la importancia de eliminar la hipótesis de campo medio en zonas donde la movilidad y las agrupaciones de grandes números de personas son tan relevantes, que es la razón por la que estamos utilizando estos modelos en primer lugar.

Creemos que este formalismo es lo suficientemente preciso y realista como para poder empezar a aplicarlo a situaciones reales, ya que está listo para añadir modificaciones como el efecto de aplicar restricciones de movilidad, cuarentena, distanciamiento social, e incluso vacunas o pérdida de inmunidad, que son clave a la hora de evaluar las respuestas de diferentes países ante la aparición de brotes epidémicos.



## Referencias

- [1] N. Boccara, *Modeling complex systems*. Springer Science & Business Media, 2010.
- [2] S. H. Strogatz, *Sync: How order emerges from chaos in the universe, nature, and daily life*. Hachette UK, 2012.
- [3] M. Resnick, *Turtles, termites, and traffic jams: Explorations in massively parallel micro-worlds*. Mit Press, 1997.
- [4] B. Gonçalves and N. Perra, *Social phenomena: From data analysis to models*. Springer, 2015.
- [5] W. Souma, Y. Fujiwara, and H. Aoyama, “Complex networks and economics,” *Physica A: Statistical Mechanics and its Applications*, vol. 324, no. 1-2, pp. 396–401, 2003.
- [6] N. Biggs, E. K. Lloyd, and R. J. Wilson, *Graph Theory, 1736-1936*. Oxford University Press, 1986.
- [7] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, “Epidemic processes in complex networks,” *Reviews of modern physics*, vol. 87, no. 3, p. 925, 2015.
- [8] D. Brockmann, L. Hufnagel, and T. Geisel, “The scaling laws of human travel,” *Nature*, vol. 439, no. 7075, pp. 462–465, 2006.
- [9] I. Rhee, M. Shin, S. Hong, K. Lee, S. J. Kim, and S. Chong, “On the levy-walk nature of human mobility,” *IEEE/ACM transactions on networking*, vol. 19, no. 3, pp. 630–643, 2011.
- [10] W. O. Kermack and A. G. McKendrick, “A contribution to the mathematical theory of epidemics,” *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, vol. 115, no. 772, pp. 700–721, 1927.
- [11] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008.
- [12] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, “Complex networks: Structure and dynamics,” *Physics Reports*, vol. 424, no. 4, pp. 175–308, 2006. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S037015730500462X>
- [13] P. Erdos and A. Renyi, “On the evolution of random graphs,” *Publ. Math. Inst. Hungary. Acad. Sci.*, vol. 5, pp. 17–61, 1960.
- [14] A.-L. Barabási and R. Albert, “Emergence of scaling in random networks,” *Science*, vol. 286, no. 5439, pp. 509–512, 1999. [Online]. Available: <https://science.sciencemag.org/content/286/5439/509>
- [15] L. Sattenspiel and K. Dietz, “A structured epidemic model incorporating geographic mobility among regions,” *Mathematical Biosciences*, vol. 128, no. 1, pp. 71–91, 1995. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/002555649400068B>
- [16] C. Cosner, J. C. Beier, R. S. Cantrell, D. Impoinvil, L. Kapitanski, M. D. Potts, A. Troyo, and S. Ruan, “The effects of human movement on the persistence of vector-borne diseases,” *Journal of theoretical biology*, vol. 258, no. 4, pp. 550–560, 2009.
- [17] D. T. Citron, C. A. Guerra, A. J. Dolgert, S. L. Wu, J. M. Henry, H. M. Sánchez C., and D. L. Smith, “Comparing metapopulation dynamics of infectious diseases under different models of human movement,” *Proceedings of the National Academy of Sciences*, vol. 118, no. 18, 2021. [Online]. Available: <https://www.pnas.org/content/118/18/e2007488118>

- [18] J. Gómez-Gardeñes, D. Soriano-Paños, and A. Arenas, “Critical regimes driven by recurrent mobility patterns of reaction–diffusion processes in networks,” *Nature Physics*, vol. 14, no. 4, pp. 391–395, 2018.
- [19] D. Soriano-Paños, L. Lotero, A. Arenas, and J. Gómez-Gardeñes, “Spreading processes in multiplex metapopulations containing different mobility networks,” *Phys. Rev. X*, vol. 8, p. 031039, Aug 2018. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevX.8.031039>
- [20] C. Granell and P. J. Mucha, “Epidemic spreading in localized environments with recurrent mobility patterns,” *Phys. Rev. E*, vol. 97, p. 052302, May 2018. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevE.97.052302>
- [21] A. Arenas, W. Cota, J. Gómez-Gardeñes, S. Gómez, C. Granell, J. T. Matamalas, D. Soriano-Paños, and B. Steinegger, “Modeling the spatiotemporal epidemic spreading of covid-19 and the impact of mobility and social distancing interventions,” *Phys. Rev. X*, vol. 10, p. 041055, Dec 2020. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevX.10.041055>
- [22] H. Hu, K. Nigmatulina, and P. Eckhoff, “The scaling of contact rates with population density for the infectious disease models,” *Mathematical Biosciences*, vol. 244, no. 2, pp. 125–134, 2013. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0025556413001235>
- [23] K. Prem, A. R. Cook, and M. Jit, “Projecting social contact matrices in 152 countries using contact surveys and demographic data,” *PLoS computational biology*, vol. 13, no. 9, 2017.
- [24] O. Diekmann, J. Heesterbeek, and M. G. Roberts, “The construction of next-generation matrices for compartmental epidemic models,” *Journal of the Royal Society Interface*, vol. 7, no. 47, pp. 873–885, 2010. [Online]. Available: <https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2009.0386>
- [25] B. J. Broxson, “The kronecker product,” *UNF Graduate Theses and Dissertations*, 2006. [Online]. Available: <https://digitalcommons.unf.edu/etd/25>
- [26] H. Hu, K. Nigmatulina, and P. Eckhoff, “The scaling of contact rates with population density for the infectious disease models,” *Mathematical Biosciences*, vol. 244, no. 2, pp. 125–134, 2013. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0025556413001235>